

AID202

Fundamentals of Machine Learning

Lecture 1:
Course outline and logistics,
An overview of Machine Learning

Today's Schedule

- Course outline and logistics
- An overview of Machine Learning

Course outline and logistics

Logistics

- **Instructor:**



Erkut ERDEM

erkut@cs.hacettepe.edu.tr

- **Teaching Assistants:** TBA
- **Lectures:** Tue 12:40 - 15:30 @D10
- **Practicum:** Tue 15:40 - 17:30 @D10

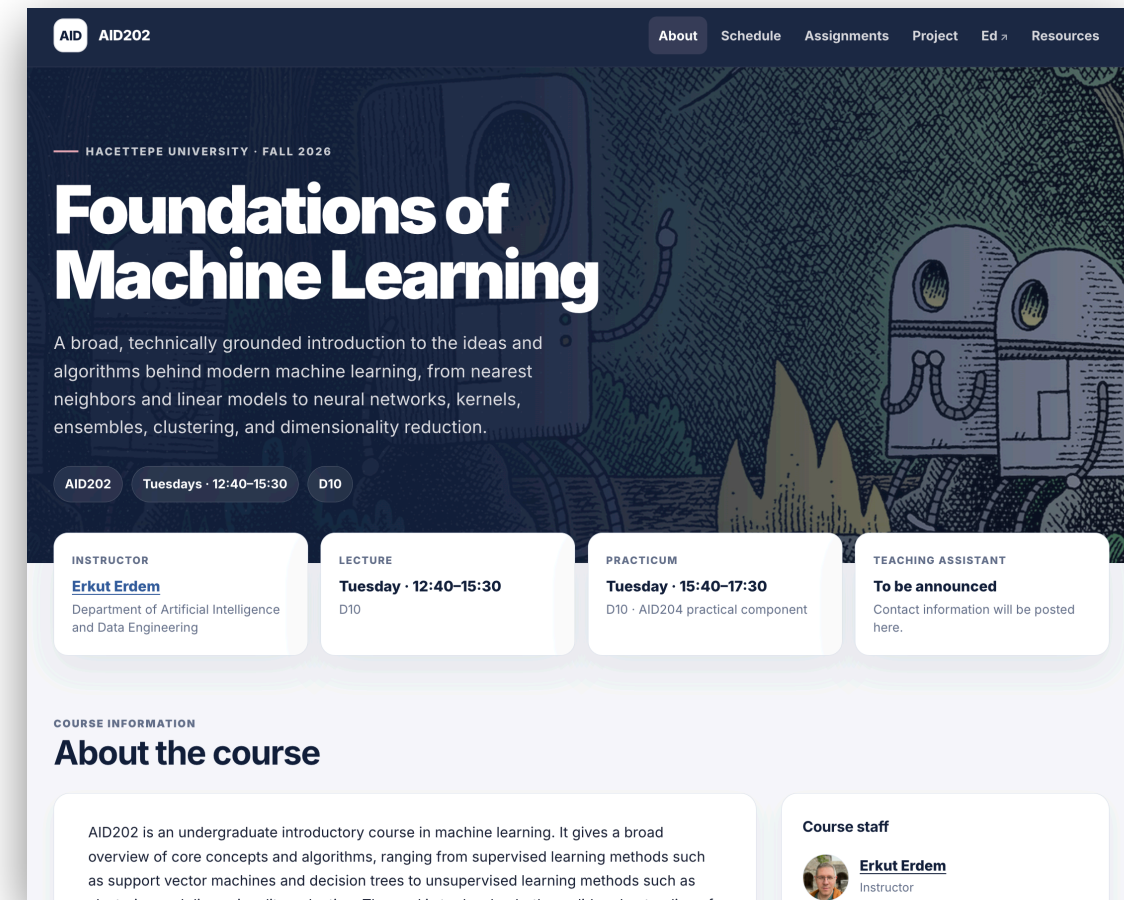
About this course

- This is a undergraduate-level introductory course in machine learning (ML)
 - A broad overview of many concepts and algorithms in ML.
- **Requirements**
 - Basic algorithms, data structures.
 - Basic probability and statistics. common distributions, Bayes rule, mean/median/mode
 - Basic linear algebra and calculus vector/matrix manipulations, partial derivatives
 - Good programming skills
- **AIN 202 Machine Learning Laboratory**
 - Students will gain skills to apply the concepts to real world problems.

Communication

- **Course webpage:**
<https://web.cs.hacettepe.edu.tr/~erkut/aid202.f26>

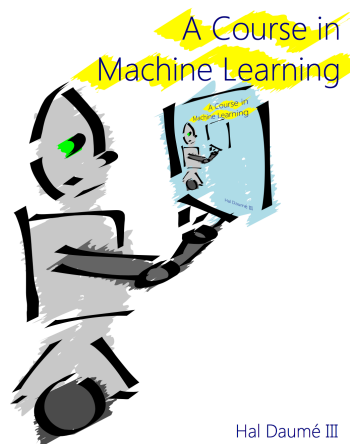
- The course webpage will be updated regularly throughout the semester with lecture notes, programming and reading assignments and important deadlines.



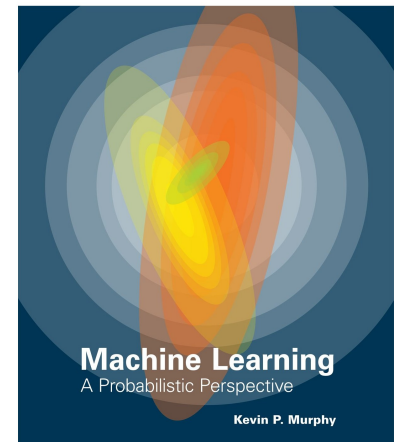
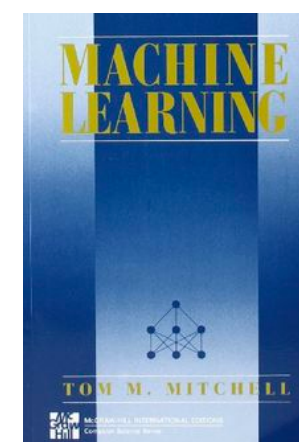
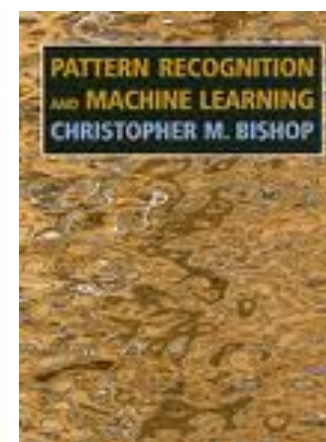
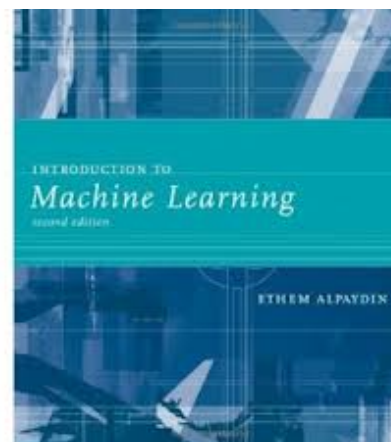
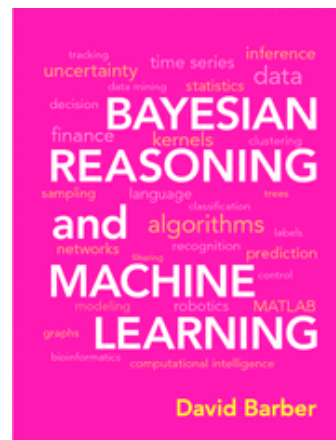
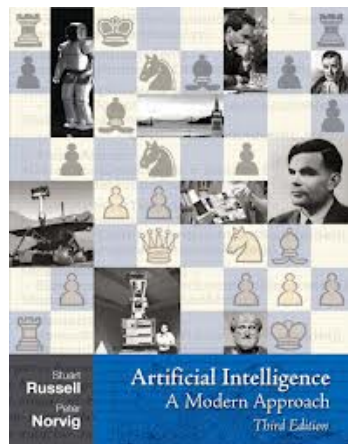
- We will be using eds for course related discussions and announcements. Please enroll the class on ed by following the link
<https://edstem.org/eu/join/7g9t3j>

Reference Books

- A Course in Machine Learning, Hal Daumé III, 2017 (**available online**)
- Artificial Intelligence: A Modern Approach (3rd Edition), Russell and Norvig. Prentice Hall, 2009
- Bayesian Reasoning and Machine Learning, Barber, Cambridge University Press, 2012 (**available online**)
- Introduction to Machine Learning (2nd Edition), Alpaydin, MIT Press, 2010
- Pattern Recognition and Machine Learning, Christopher Bishop, Springer, 2006 (**available online**)
- Machine Learning, Tom Mitchell, McGraw Hill, 1997 (**available online**)
- Machine Learning: A Probabilistic Perspective, Murphy, MIT Press, 2012



Hal Daumé III



Grading Policy

- Grading for AID 202 will be based on
 - class participation (6%)
 - course project (**done in groups of 2-3 students**) (35%),
 - midterm exam (27%), and
 - final exam (32%)
- In AID 204, the grading will be based on
 - a set of quizzes (20%) (**the lowest quiz grade will be dropped**), and
 - 3 assignments (80%) (**done individually**)

Assignments

- 3 assignments
 - First one worths 20%, last two worth 30% each
- **Theoretical:** Pencil-and-paper derivations
- **Programming:** Implementing Python code to solve a given real-world problem
- **Late policy:** You have 5 slip days in the semester.



**KEEP
CALM
AND
DO YOUR
HOMEWORKS**

Course Project

- Done in groups of 2-3 students.
- Choose your own topic (but focused on a specific theme) and explore ways to solve the problem
- This year's theme is will be announced soon.
- **Proposal:** 1 page (Oct 23) (3%)
- **Progress Report:** 5 pages (Dec 8) (10%)
- **Project Presentation:** Classroom presentation (Dec 22) (10%)
- **Final Report:** 9 pages (Dec 31) (12%)

Sample projects from 2016 (BBM406)

— Food and ML —

BBM 406 Class Project - Final Report

Cem Güngör, Fatih Baltacı
Department of Computer Engineering
Hacettepe University
Ankara - TURKEY, Fall 2016
{b21328031, b21327689}@cs.hacettepe.edu.tr



Abstract

This paper is a final report of our project "What Am I Eating?" for BBM406 Introduction to Machine Learning lesson. "What Am I Eating?" is an image recognition project which predicts food labels from given images. Developments in the field of Machine Learning and increase of datasets in recent years encourage us to make an image recognition project. We are using deep learning. We performed transfer learning (from Inception v3 model [Szegedy et al. 2015]) and data augmentation. Our dataset is a combination of different datasets which has 113 classes. Each class has 1000 images.

Keywords: deep learning, image recognition, fine tuning

1 Introduction

In recent years there have been major developments in the field of machine learning. The datasets have grown up because of the increase in internet usage. Hardwares become stronger than before. Graphic cards become cheaper. Because of these conditions, researches have increased and new approaches such as deep learning has appeared. Open source libraries were developed.

Deep Learning is a new and very popular area of Machine Learning research. We decided to develop a project using deep learning to improve ourselves in this field. Deep learning is used in many

areas such as image recognition, speech recognition, natural language processing and so on. We used deep learning for image recognition. So, What am I Eating? is a deep learning project that recognizes foods from images.

We saw that no dataset has any Turkish foods. We wanted our project to recognize Turkish foods too. Also we have some future thoughts about our project.



Figure 1: pizza (score = 0.84349), waffle (score = 0.04952), bruschetta (score = 0.02402), omelette (score = 0.01936), ...

2 Related Work

There are three researches which are closely related to our research topic. All of them are new and made in 2016. One of them is [Liu et al. 2016]. The purpose of this research is to improve the accuracy of current measurements of dietary intake by analyzing the food images captured by mobile

PREDICTING RESTAURANT RATINGS FROM REVIEW TEXTS

Melis Mutlu & Raisha Abdillab
Department of Computer Engineering
Hacettepe University
Beytepe, Ankara 06800, Turkey
{b36160843, b21304258}@cs.hacettepe.edu.tr

ABSTRACT

Nowadays, with the growth of crowd-sourced review website, the quality of business is determined by its ratings and reviews. The customer and the business owner will be able to see the trends, making decision, and getting recommendations based on their preferences just by looking at the reviews and ratings themselves. In this project, our goal is to predict the ratings which is given to a restaurant by looking at its review text. We use Yelp Dataset for our training and testing. By applying machine learning and text mining principle, we analyzed the review text from the Yelp Dataset. We were researching for the best algorithm which would give us the best result. The algorithms which we used at this projects are Bayesian Ridge Regression, Support Vector Regression, and Random Forest Regression.

1 INTRODUCTION

The development of technology makes it easier for people to make the right decisions. In this matter, technology influences the field of business by delivering a more convenient way for people to evaluate their business. For example, nowadays customer may look at the reviews and ratings which has been given and getting influenced by it, before deciding to go to a certain restaurant.

The goal of our project is to choose a supervised machine learning algorithm which will give us the best performance in predicting the restaurant ratings by looking at its review text that has been given in Yelp Dataset. Firstly we have to choose the most appropriate dataset to our problem. After that, in order to work with Machine Learning algorithm, we transform our raw data into vector or matrices form.

For our project we use Yelp Dataset, since it already provides the review and rating in an easily accessible format. Then, we did feature extraction from our dataset. We combined several feature extracting process in order to get the better result. For this, we use Bag of Words and Word2Vec model. We have tested these model and it gave us a satisfying result. For the better result, we also removed words which we considered unimportant. After we made our model, we use machine learning algorithm to test our model. We then choose the algorithm which gave us the best performance after we tested it. We treated this problem as regression problem, therefore we used regression algorithm. We made use of Yelp Dataset as our training set and testing set.

In this report, firstly, we will present you the dataset. Secondly, we will tell you about our feature extraction method (Bag of Words, Word2Vec). The next part is that we will explain about the algorithm which we use for this projects, which consists of Bayesian Ridge Regression, Support Vector Regression and Random Forest Regression. Then, by using Explained Variance Score (R^2 score) and Mean Square Error we calculate the accuracy of our model. We will share the result and the conclusion of our project by the last part of this report.

1

Finding The Ingredients of Pizza Using Deep Learning

Mümin Can Yılmaz Alim Giray Aytar
can.yilmaz12@hacettepe.edu.tr giray.aytar12@hacettepe.edu.tr

Hayati İbiş
hayati.ibis12@hacettepe.edu.tr

Abstract

Extracting ingredients from a dish can be a powerful tool for combatting obesity and making food inspection processes easier. For this purpose, we tried to create a program which extracts ingredients from a pizza, using convolutional neural networks. We also created a dataset which has 7405 images and 20 different labels as ingredients. Our experiments show us our model can predict small numbers of ingredients successfully (80 percent for one label), however as the number of ingredients increased, accuracy rate drops significantly (22 percent for 2 labels).

1. Introduction

Our aim is to create a model which can identify ingredients in the pizza. Our program should output a list of ingredients as output when feed with an image of a pizza.

First of all, we started with creating a new dataset from the scratch, because we couldn't find any ready-to-use dataset. To do this, we collected about twenty five thousand images from web and labeled all of them by hand with a little software we created for this purpose.

Secondly, we decided to use a Convolutional Neural Network, because they show much better performance in image recognition problems compared to other approaches. Also when using Convolutional Neural Networks, we don't need to extract any features because CNN's operates directly on images. There is also some downsides of using Convolutional Neural Networks as they need more data and require more computing power than other solutions.

Finally, we evaluated our project with the result that we get after the process of training our classifier model which we present in the results section.

Hardest part of this problem is, because food shapes are deformed after cooking, it might not be possible to predict them correctly for our model. Color information also isn't very helpful, because some different ingredients exactly have the same colour or same ingredients might have different colours.

1



hamsi: 0.58653
baklava: 0.30801
carrot cake: 0.05741
humus: 0.01253



```
{
  'type': 'business',
  'business_id': (encrypted business id),
  'name': (business name),
  'neighborhoods': [(hood names)],
  'full_address': (localized address),
  'city': (city),
  'state': (state),
  'latitude': latitude,
  'longitude': longitude,
  'review_count': review count,
  'categories': [(localized category names)]
  'open': True / False (corresponds to permanently closed, not business hours),
}
```



Green Pepper
Olive
Onion
Salami
Corn
Chicken
....

Sample projects from 2017 (BBM406)

— City and ML —

Predicting the Location of a Photograph

Ali Yunus Emre ÖZKÖSE
Hacettepe University
ANKARA, TURKEY
aliozkose@hacettepe.edu.tr

Tarik Ayberk YILIKOĞLU
Hacettepe University
ANKARA, TURKEY
tarikyilkoglu@hacettepe.edu.tr

Abstract

In this paper, we addressed to prediction of an image location problem. It is still a hard problem because of several kinds of other problems. We use convolutional neural networks (CNNs) to tackle this problem. We collect data from Flickr[13], create a dataset which we call Turkey15 and test with basic algorithms. After testing the dataset, we train AlexNet and ResNet-18 with Turkey15 from scratch. Since Turkey15 is very small, we use transfer learning to improve results. We use feature extracting and fine-tuning[14]. We also freeze some layers to get better accuracy.



Figure 1: Images from Turkey15

1. Introduction

Although there are a lot of works on this issue and it is very popular research topic in recent years, predicting the location of an image is still a hard problem. There are various problems such that constructing features [3], viewpoint problem[4], illumination and structural modification[12] etc.. It can be used for many areas such as estimation people's perception [5]. But how can we predict the location of given image? In this work, we focus on exactly the problem of city classification.

With the development of technology and the increase of applications, people are taking photos and upload to internet much more than ever. The significant point of sharing is that a huge data has existed and it can be used for creating artificial machines as an experience. At this point, we collected images from Flickr where are taken in Turkey, cre-



Figure 2: Images from Turkey15

ated a dataset which we called Turkey15 and predict image locations where is limited to Turkey.

First of all, we tested our dataset with hand-crafted features which are Tiny images, GIST features, and Hog features, because we should know that our dataset is convenient enough to use as a dataset or not. Details in this process are explained in section 3.1.

After testing the dataset, we trained existed models which are AlexNet and ResNet-18 models with our dataset. We trained from scratch in this step and get some results and compare with training with hand-crafted features. Details are written in section 4.1.

Thirdly, we used transfer learning, in particular, fine-tuning and feature extracting. We trained pre-trained models which are trained with places365 and imageNet datasets. Models are AlexNet and ResNet-18 again. Details are written in section 4.2.

Finally, we froze some layers of models and trained AlexNet and ResNet-18 again. Details are written in section 4.3.

2. Related Work

Because of the popularity of this challenge, there are many kinds of proposed methods and works for predicting location. Li et al. propose to represent features with SIFT and match query image features to database image features mutually[11], but matching is only among the prioritized features. They keep informative points. In this way, they reduce computational cost. We also used hand-crafted features for testing dataset, but we use convolutional neural networks for training.

Sound of The City

Buğrahan Akbulut
Department of CS Hacettepe University
akbulutbugrahan@gmail.com

Mustafa Çağdaş Çaylı
Department of CS Hacettepe University
cagdas_cayli@hotmail.com

Abstract

In this paper we will introduce our project that is detects and classify leading sounds on urban sounds. We focused on audio because it was more attractive then working on image or some numerical data and also because sound is a very important tool for understanding the world. Also another reason is working with sound is very challenging because it is hard to find only one pure sound on outside world there are lots of sound sources and we generally hear the mixture of these sounds, so our data sets that we used in this project have real field records - has lots of mixed sounds. We worked on UrbanSound8K and UrbanSound data sets containing 27 hours of audio with 18.5 hours of annotated sound event occurrences across 10 sound classes (air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren, and street music). Our goal was extract leading sounds with a correct shape by using Shogun and classify them correctly.

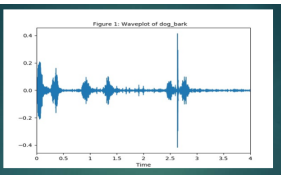
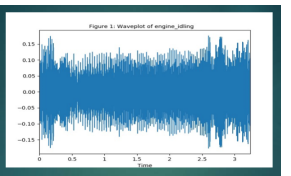
1. Introduction

Since new audio technologies developed rapidly recently, audio processing and classification are growing research fields and it contains many challenges. Especially separating audio into its components is a very tough problem. However working on an analysis of urban sounds instead of working on the analysis of speech, music, bioacoustics is relatively easy and relaxing. Furthermore we worked on extraction of the leading sounds with correct shape.

One of the main challenges in this project was lack of labeled mixed sounds. Previous work focused on classification of single labeled audio data. We needed lots of audio data to get our final results correct. With this purpose we created our own multi-labeled audios by using Shogun. Actually we first wanted to separate a given any kind of mixed sound into its components by using ICA (independent component analysis) but we could not find any working library or implementation of this algorithm and due to the restricted time we could not achieve this

goal. But we wanted to make it so we have done some more researches and find a new library named shogun which provide some tools for mixing and separating sounds not like ICA but it works for us to get some results by making tests on mixed sounds.

After all these things we also want to improve our results getting from tests, we decided to combine two different machine learning approach to get higher results and it was another challenge for us to increase our results by using neural networks and support vector machines combination. The approach we use to combine these two algorithms will be explained in more detail at "The Approach" section. Here you can see wave-plot form of single and mixed sound sources we worked on :



Prediction Of Life Quality

Tark Ramazan BASOGLU
Department of Computer Engineering
Hacettepe University, Ankara, TURKEY
tarik.basoglu@hacettepe.edu.tr

Emre DOĞAN
Department of Computer Engineering
Hacettepe University, Ankara, TURKEY
emre.dogan@hacettepe.edu.tr

Abstract

In this study, we mention about the usage of using a machine learning approach to specify life qualities of cities instead of public research. We create an assorted dataset that contains statistical and physical features. To do that, we utilize from MAPZEN. We expect to predict the scores on MOVEHUB with high accuracy.

1. Introduction

Nowadays, we can easily see that cities differ considerably from each other in terms of their physical and social characteristics and that difference is highly influential in human life. We are making great efforts to determine the effects of these differences on human life and to make cities more livable and to change this imbalance positively.

In this situation, we are faced with a notion named quality of life.

"Quality of life (QOL) is the general well-being of individuals and societies, outlining negative and positive features of life. It observes life satisfaction, including everything from physical health, family, education, employment, wealth, religious beliefs, finance and the environment." [2]

By this definition, there are various social and physical criteria that influence the quality of life. The number of researches and studies carried out in this area is increasing day by day. While life quality information for large cities is easily accessible, it is not possible to find reliable results for cities that are not big enough.

In this project, we purpose to achieve higher efficiency in shorter time and reduce the burden on a human in such researches. Rather the laborious and time-consuming processes of public researches we also aim to provide a new, flexible and developable method by making use of

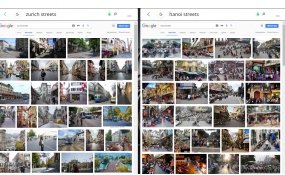


Figure 1: The reflection of the crowd difference between the Hanoi and Zurich on the street photos

machine learning experiences. Thus, we get a chance to detect the life qualities for any cities in the world. At the same time, we are expecting to be able to observe which physical factors effects the life quality with which rates.

MoVEHUB

There is a platform named MOVEHUB that helps you make informed decisions about where to move to around the world. And it has a city ranking list consists of over 200 cities. We utilized this list as the main target in the estimation results.

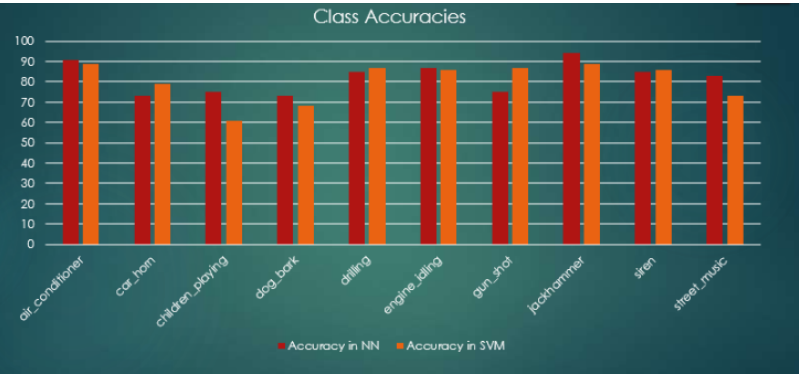
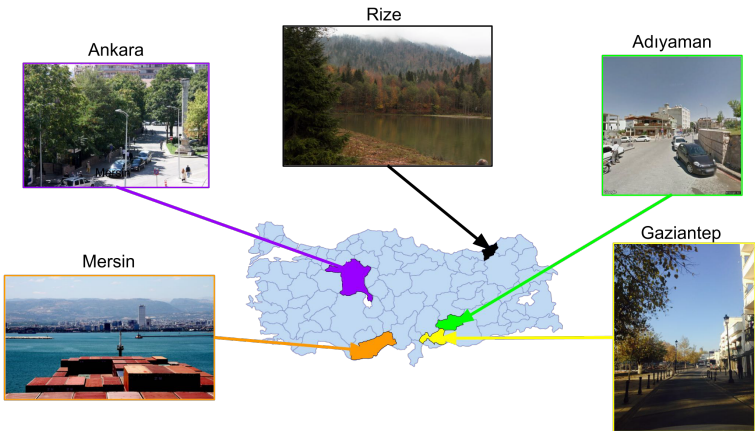
MAPZEN

Mapzen is an open and accessible mapping platform that is focused on the core components of geo platforms, including search, rendering, navigation, and data.

2. Related Work

There are numberless researchs done to measure life quality in cities every year. In this researches generally, lots of criteria are considered to obtain correct results. Such researches have been carried out in the form of public opinion polls up to now.

MOVEHUB: MOVEHUB is similar research that includes



```
landusages.geojson
{"id": 2, "osm_id": -156806, "name": null, "type": "forest", "area": 0.000004, "z_order": 45, "population": 0},
{"id": 3, "osm_id": -319722, "name": null, "type": "pedestrian", "area": 0.000000, "z_order": 49, "population": 0},
{"id": 67, "osm_id": -2562662, "name": null, "type": "farmland", "area": 0.000004, "z_order": 41, "population": 0},
{"id": 131, "osm_id": -444236, "name": null, "type": "park", "area": 0.000001, "z_order": 46, "population": 0},
{"id": 132, "osm_id": -5305314, "name": null, "type": "scrub", "area": 0.000001, "z_order": 16, "population": 0},
{"id": 208, "osm_id": 11694091, "name": null, "type": "parking", "area": 0.000000, "z_order": 22, "population": 0},
{"id": 209, "osm_id": 12000065, "name": null, "type": "playground", "area": 0.000000, "z_order": 47, "population": 0},
places.geojson
{"id": 14, "osm_id": 91115140, "name": "Dorff", "type": "village", "z_order": 5, "population": 0},
{"id": 15, "osm_id": 97038695, "name": "Verlautenheide", "type": "suburb", "z_order": 3, "population": 0},
{"id": 22, "osm_id": 240041315, "name": "Aachen", "type": "city", "z_order": 7, "population": 0},
amenities.geojson
{"id": 56, "osm_id": 3657599191, "name": "Brandweer Bocholtz", "type": "fire_station", "z_order": 4, "population": 0},
{"id": 58, "osm_id": 3741421875, "name": "B77cherinsel St. Donatus", "type": "library", "z_order": 4, "population": 0},
{"id": 59, "osm_id": 4205246567, "name": "Bibliothek Katho", "type": "university", "z_order": 4, "population": 0},
{"id": 60, "osm_id": 4501507309, "name": "Shell Tankstelle", "type": "fuel", "z_order": 4, "population": 0},
buildings.geojson
{"id": 2, "osm_id": -361947, "name": null, "type": "house", "z_order": 1, "population": 0},
{"id": 5, "osm_id": -532049, "name": null, "type": "commercial", "z_order": 1, "population": 0},
{"id": 9, "osm_id": -898929, "name": null, "type": "industrial", "z_order": 1, "population": 0},
{"id": 10, "osm_id": -898909, "name": "Zorgcentrum De Schutse", "type": "apartments", "z_order": 1, "population": 0},
{"id": 11, "osm_id": -898915, "name": null, "type": "house", "z_order": 1, "population": 0},
roads.geojson
{"id": 99, "osm_id": 4741398, "name": "cycleway", "type": null, "tunnel": 0, "bridge": 0, "oneway": 0},
{"id": 100, "osm_id": 4754226, "name": "footway", "type": null, "tunnel": 0, "bridge": 0, "oneway": 0},
{"id": 101, "osm_id": 4754228, "name": "pedestrian", "type": null, "tunnel": 0, "bridge": 0, "oneway": 0},
{"id": 434, "osm_id": 6585093, "name": "living_street", "type": "Kerkbuurt", "tunnel": 0, "bridge": 0, "oneway": 0}
```

Sample projects from 2018 (BBM406)

— Home and ML —

Wi-Fi Based Indoor Positioning Systems

Burak Emre Ozer
Hacettepe University
Ankara, Turkey
burak.ozer@hacettepe.edu.tr

Furkan Caglar Gulmez
Hacettepe University
Ankara, Turkey
furkancaglargulmez@gmail.com

Huzeyfe Kocabas
Hacettepe University
Ankara, Turkey
kocabashuzeyfe@gmail.com



Figure 1: Source [12] Left: map of UII campus and Tx buildings. Middle: red indicates ESTCE - Tx building. Right: example of a reference point.

Abstract

Wi-Fi Fingerprint-based positioning approach that detects the position of user or device is widely used in the indoor positioning systems instead of Global Positioning System (GPS). In this approach, Received Signal Strength (RSS) values that are known as Wi-Fi fingerprints used. Received Signal Strength values are the measurement of the power present in a received radio signals. We use UIIIndoorLoc dataset with 19937 training records and 1111 test records. This dataset of RSS values are collected by using previously placed wireless access points (WPs) in Tx Buildings of University of Jaume I campus. We aim to predict location points with respect to floor IDs, building IDs, longitude and latitude values with supervised machine learning algorithms such as K-Nearest Neighbor Algorithm, Random Forest Algorithm, Support Vector Machine and Decision Tree Algorithm. Then we use the model with the highest accuracy in the rest of the progress. Classification techniques are used for building and floor classification and regression techniques are used for detection of location points.

Keywords— indoor positioning, received signal strength indication (RSSI), machine learning algorithms, classification, random forest

Country Classification Using House Photos

Meltem TOKGOZ
Hacettepe University
21527381
meltemtokgoz@hacettepe.edu.tr

Enes Furkan CIGDEM
Hacettepe University
21526877
enescigdem@hacettepe.edu.tr

Asma AIOUEZ
Hacettepe University
21504074
asma.aiouvez@hacettepe.edu.tr

Abstract

Home designs vary from country to country and when we talk about housing, we should refer to both modern and traditional styles. You can come across a picture of a house taken by someone anywhere in the world and you may wonder where it has been taken from. In this project, we tried to find out which country the photo of a house was taken from. In short, we worked on the problem of classification according to where the photographs were taken.

We used our own World dataset for this project. This dataset contains over 4000 pictures for 15 different countries. In our project, we collected our data from the Flickr [1], Pinterest [3], and Google Photos [2]. We first tested our data with a single layer neural network and then with convolutional neural networks (CNN). We used ResNet18 and AlexNet models when implementing CNN in our project. In accordance with the results, we applied some methods to increase the accuracy and we got the best accuracy with ResNet18.

1. Introduction

Recognizing home photos and classifying them by country is a quite difficult problem. Because the houses in many countries in the modern world are similar to each other. Beside that, there are some features to distinguish these houses. For example, each country's climate, people's lifestyle and culture are different. This gives us some hints on the architecture of the houses in that country. From this point of view, especially the design of traditionally styled houses begins to change from a country to another. The main problem here is that the houses in the same continent are very similar to each other. For example as shown in

Figure 1, in the Asian continent, traditionally styled houses of some countries such as South Korea, Japan, Indonesia and Malaysia are very similar. This factor complicates the solution of the problem. In addition, many factors such as the shooting angle, light, shadow and seasonal differences affect the solution of this problem.



Figure 1: Example of similar data

Since this is an image classification problem, there are many algorithms and methods used in its solution. K-nearest neighbors, logistic regression, support vector machine and convolutional neural networks are some of these solutions. Especially in recent years, CNN is a successful algorithm preferred to solving problems in this area.



Figure 2: Example of similar data

In our study, we deal with the problem of classification according to the country where the house pictures were

Rock or Not?

Defne Tuncer¹ Kutay Barcin¹

Abstract

In the era of technology, millions of songs are brought to people everyday. The dramatic increase in the size of music collections has made the music genre recognition (MGR) an important task on machine learning. The goal of this paper is to give machines a chance to predict music genres given input features from music tracks. To do that, we applied various techniques based on machine learning on the dataset called Free Music Archive (FMA), and we have reached an accuracy score of 67.80% as our highest.

solvers and regularization. 4.1.3 Support Vector Machines with linear and radial basis function (RBF) kernels. 4.1.4 Deep Learning method Neural Network also known as Multi-Layer Perceptron through various optimizers. To represent the audio tracks in building our baseline models we planned to use the combination of all the features, which have been shown to be effective in the task of predicting genres. We improved our methods with model and feature selection by using k-fold cross validation afterwards. Based on the results obtained from the algorithms, we performed experimental analysis. Finally, ended our work with a detailed conclusion, and proposed our feature work.

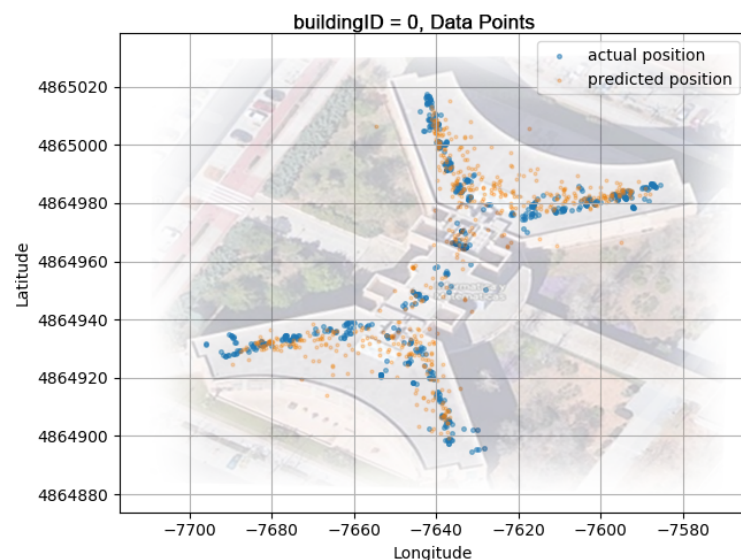
2. Related Work

For the music genre recognition task, the most common datasets are GTZAN (Tzanetakis & Cook, 2002), Million Song Dataset (MSD) (Bertin-mahieux et al., 2011) and FMA: A Dataset For Music Analysis (Defferrard et al., 2017). While FMA, which consists of 161 sub-genres among 106,574 tracks and published in 2017, is the most up-to-date dataset, and is especially suited for MGR as it features fine genre information. A challenge took place as one of challenges of Web Conference (WWW2018) by the publishers of FMA Dataset on the subject predicting genres of the music (Defferrard et al., 2018). The winner succeeded by examining through artist-related information and scored an accuracy of 66.29% on predicting 16 genres (Kim et al., 2018).

In Music Information Retrieval (MIR), there have been various number of studies on building effective models to predict genre of music using audio features. Mel-Frequency Cepstral Coefficients (MFCCs), one of the audio features, are generally used in music genre classification as the perceptual scale of pitches of a human hearing are represented by the Mel-scale. A Hidden Markov model with MFCCs is used to classify pop, country, jazz and classical genres in (Shao et al., 2004). On the other hand, another study focuses on a new feature called Renyi Entropy Cepstral Coefficients (RECCs) (Tsai & Bao, 2010). The highest achieved accuracy scores reported on the datasets ISMIR2004 which is from the contest (Cano et al., 2006) and GTZAN are accomplished by representing the auditory human perception with a proposed spectrogram (Panagakis et al., 2009). Most of their studies are done through researching the timbre texture,

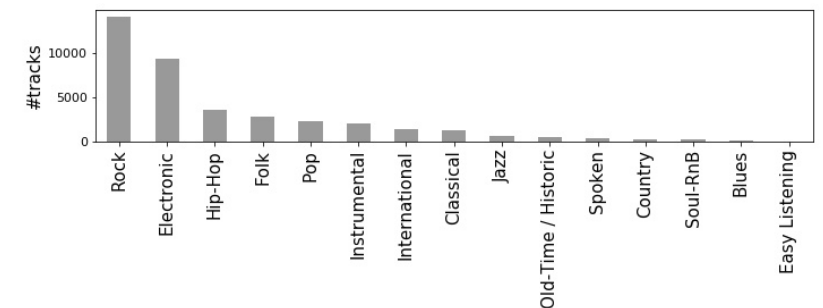
¹Department of Computer Engineering, Hacettepe University, Ankara, Turkey. Correspondence to: Defne Tuncer <defnetuncer@hacettepe.edu.tr>, Kutay Barcin <kutaybarcin@hacettepe.edu.tr>.

BBM406 Fundamentals of Machine Learning, Fall 2018. Copyright 2018 by the author(s).



Predict Class : Indonesia

Correct Class : Malaysia



Sample projects from 2019 (BBM406)

— ML and Social Good —

ClearSky: Satellite Imagery Based Deep Learning Model to Predict Wildfire Spread

Ahmet Ş. Yener Deniz Zağlı

Abstract

As wildfires rage across the entire continent of Australia, the lack of resources are more apparent than ever. When there are limited resources the lack of forecasts make the distribution of resources inefficient. This inefficiency has a wide range of effects, from human and animal life losses to monetary damage. Current in use forecast systems require expert knowledge to run and are case to case basis so the data available is not leveraged to the extent it should. We've developed a system that alleviates these problems combining GIS (Geographic Information Systems) and Deep Learning models that are used in meaning extraction from visual data. We've expanded and showed the agnosticism of the model to spatial,temporal data and underlying code frameworks. We propose future work that will increase efficiency and automate the system, making it distributable to the authorities around the globe.

1. Introduction

The current climate change effects create much more suitable conditions for wildfire outbreaks and spread (Clarke, Hamish, et al. (2019)). Wildfires across the globe are expected to rise in amount and intensity (Running, Steven W.,James, Sagil, et al. "Smart Drone Technology for Wildfire Prediction and Prevention," (2019)., (Chung, M., M. Jung, and Y. Kim. et al. (2019).) While detection is possible, forecast on the spread of the wildfire is a critical information for minimizing the damage caused.

Data we are using is widely spread and do not exist in an analyzable form in any single place. This obscurity by diffusion limits the amount of work done on such a critical area. Data in it's provided form needs heavy processing and geo-referencing to be able to analyzed by the Deep Learning models. These obstacles can be automated with collaboration by the data providers. The upside is that the datasets used are all publicly available.

We use Satellite generated imagery combined with GIS applications to generate input and create ground-truth for the supervised Deep Learning model. The prediction ability of

Heart Disease Detection

Oğuzhan Eroğlu¹ Harun Alperen Toktaş¹

Abstract

In this paper we introduce methods for heart disease detection and data sets analysis. We experiment with three different data sets. The first data set includes 303 records and 13 features + binary target. The second data set includes over 4000 records and 14 features + binary target. The third data set includes 70,000 records of patients data, 11 features + binary target. We aim to compare different machine learning algorithms on this data sets and their results. We are going to determine which algorithm is more suitable for the data sets we use for heart disease detection problem. We are going to also compare the results we have found for all algorithms with previous studies.

1. Keywords

Heart Disease Detection, Early Diagnosis, Prediction, Classification, AI For Good.

2. Introduction

We will introduce you to the project of the machine learning course that we are taking this semester. The theme of the projects of this semester is machine learning for good. So we decided to do Heart Disease Detection. The motivation behind choosing this project is to show people that machine learning techniques can be life-saving.

We miss a lot of things while dealing with the daily hustle and bustle. The most important of these is our health. Health is better than wealth. Millions of people die every year because of sudden heart diseases. These sudden cardiac disorders may result in sudden deaths, as well as life-long disturbances reduce our quality of life?

Early diagnosis is very important in the prevention of heart diseases. But most people are unaware that they have any

Histopathological Cancer Detection by Using Convolutional Neural Network

Bahar Bender¹ Ezgi Türkokuşlu¹ Furkan Kaya¹

Abstract

According to research, the risk of people getting cancer increases with each passing year. In this case, machine learning algorithms are required to step into the health sector to provide early diagnosis and speed up the work of pathologists. This paper aims to detect malignant tumors by using histopathological data and to reveal the presence of cancer. For this purpose, an algorithm has been developed to predict the presence of metastatic cancer in small image patches (96x96px) from high-resolution pathology scans. Our data set will be a modified data set obtained by subtracting duplicates from the PatchCamelyon (PCam) benchmark dataset (Veeling et al., 2018). By supporting various parameters and improvements, we applied the Convolutional Neural Network and reached the highest AUC value of 97.56% as the measurement metric.

1. Introduction

"Cancer develops when normal cells in a particular part of the body begin to grow out of control. Cancer is the second leading cause of death in the world after cardiovascular diseases. Today, millions of cancer people extend their life due to early identification and treatment." (Sudhakar, 2009)

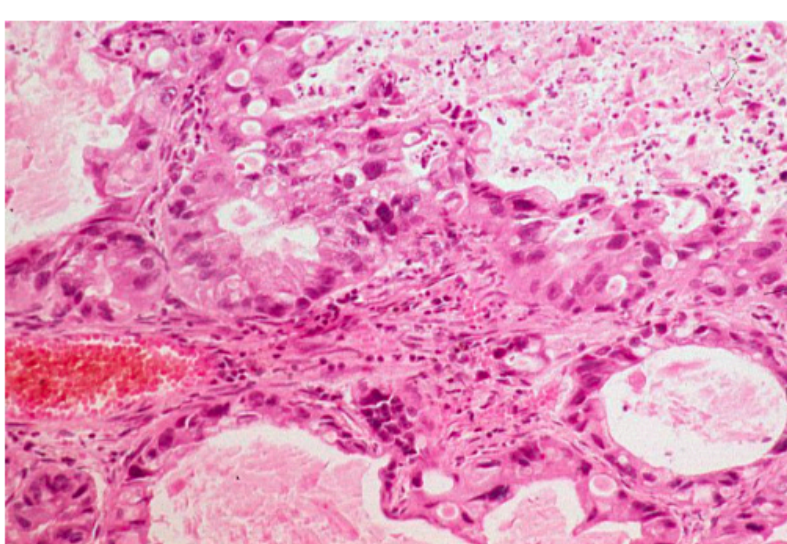
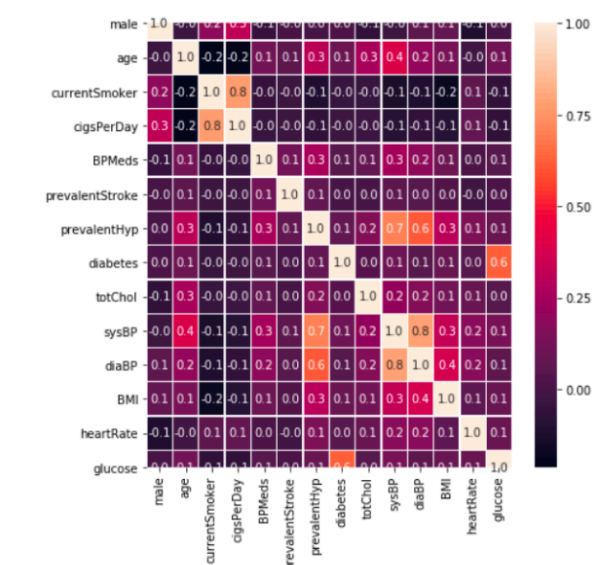
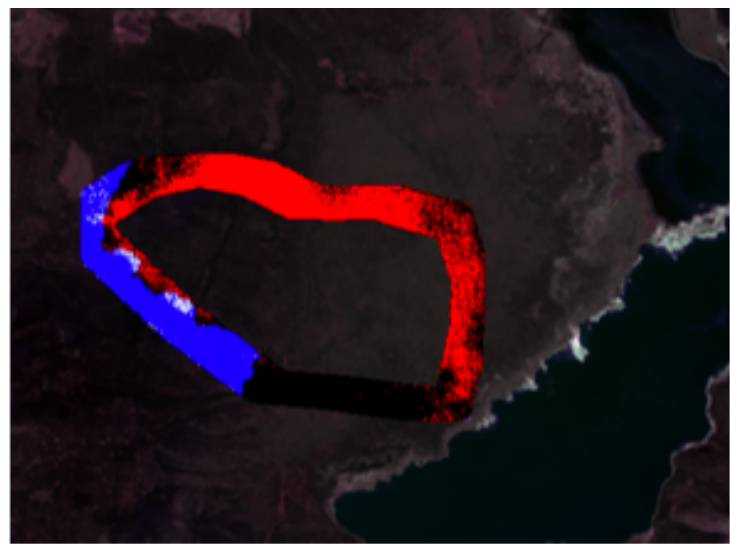
According to research, 14 million people are diagnosed with cancer every year in the world and more than 32 million people live with cancer. In our country, approximately 150 thousand cancer is diagnosed every year. Besides, cancer in almost every country is showing an increase of 1–2 percent per year. Due to advances in early diagnosis and treatment, cancer is now among the curable diseases. Therefore, the biggest motivation in the project is to be able to diagnose cancer much faster and hope for the lives of many more people.

Histopathology is a branch of pathology specializing in

Figure 1. Example of histopathological image

2. Related Work

The process of detecting cancerous tissues on histopathological images has been studied on many individuals and different classification methods and architectures have been used. According to Bastiaan S. Veeling (Xiao et al., 2017), the identification and classification of metastatic breast cancer was performed on digital full-slide images of sentinel lymph node biopsies using ResNeXt, a 101-layer network



Sample projects from Spring 2021 (BBM406)

— Art and ML —

Music Recommendation and Playlist Continuation

Arda Hüseyinoğlu

Mehmet Serkan Tan

Abstract

In this paper, recommender systems for playlist continuation are discussed. Our approach focuses on neural network-based solutions, which have yielded immense success in many fields, for playlists that are composed of very few to many songs. The proposed approach leverages and extends the Neural Collaborative Filtering model. The model makes use of embedded vectors to find similarities between playlists. Training and evaluation are done with the 1 Million Playlist dataset which was curated by Spotify for the 2018 ACM RecSys Challenge, and it's still a relevant challenge today.

1 Introduction

In information systems, recommender systems are quite important, and a useful tool in various applications to help users in identifying relevant content in a wide database. Today, most of the music consumption is done through cloud-based streaming services (e.g. Spotify, Apple Music, Tidal). The existence of a huge number of singles on these platforms challenges consumers to compile their own tastes of music. Besides, the great size and complexity of the music data space and the subjectivity of users make the problem quite hard for the companies to recommend the right music. These concerns eventually attract many researchers that are working in the information systems.

In this work, we focus on the problem of playlist continuation, which is suggesting tracks that a user may add to an existing playlist. We build a deep learning-based playlist continuation model to tackle this problem. Our main methodology is the Neural Collaborative Filtering (NCF) proposed by [1] He et al. where they use embedding layers and multi-layer perceptron to learn user-item interaction function. We use the NCF model to adapt it our problem of playlist continuation. We also aim to extend the model by providing additional input vectors such as album-id and artist-id to learn more complex relations.

This paper is organized as follows: In the second section the related work in the field is discussed, 3rd section describes the main dataset, its characteristics and the preparation procedure for training, 4th section focuses on the

learning methodology for the model, 5th section describes the evaluation techniques to measure performance, 6th section shows, and explains the experimental findings, 7th section discusses the future work, and the concluding remarks are given at the end of the paper.

2 Related Work

In 2018, when Spotify announced the playlist recommendation competition for the first time, the top works were published in the ACM's RecSys'18 at the end of the competition. There are several approaches to consider. [2] Maksim Volkovs et al. propose a two-stage technique where in the first stage the system is optimized for fast retrieval of the music data space, and in the second stage the candidates which are retrieved in the first stage are re-ranked to maximize the accuracy for the evaluation. Making use of autoencoders and character-level convolutional neural networks is an another approach[3] where they first analyze a playlist and its categorical contents and second, to make use of the playlist title. In this way, they show that popularity bias and cold start problems can be mitigated. There are also works where traditional methods are used such as [4] Sebastiano Antenucci et al. proposes a solution based on content-based and collaborative models. In addition, they boost the model on top of the final prediction by analyzing the underlying structure of the dataset. (Spotify 1 Million Playlist). They also look out for computational resources. Also, [5] Vasilij Rubtsov et al. applies collaborative filtering for the candidate selection, and gradient boosting for final prediction.

Besides the RecSys'18 publications, [1] He et al. proposed an alternative way for the collaborative filtering. They use multi-layer perceptron to learn the user-item interaction function to generalize matrix factorization under its framework. They use two feature vectors as input to describe a user and an item. Since the proposed work based on the collaborative filtering setting, identity of a user and an item are used as the input feature which is a kind of one-hot encoded vector. Embedding layers are used to project the sparse representation of the input vectors to a dense vector. Then, the user embedding and item embedding are then fed into a multi-layer neural architecture, which they are called as

Art-Style Image Transferring

Oktaý UĞURLU^{*} Koray KARA^{*} Ahmet Deniz GÜNER^{*}

Abstract

Nowadays, Generative Adversarial Networks (GANs) are an emerging technology that is used in both supervised and unsupervised learning. These networks are also capable of producing high-quality data in an efficient way. Image to image translation is one of the core applications of GANs. For instance, a data augmentation that we have used in this project. In this study, we propose the methods that keep the quality of the style transferring high. For this purpose, we are using the CycleGAN that is an extension of the GAN architecture. Generative Networks allow easy mapping between the source image and the target image. It also calculates a loss function to greatly improve the quality of this generated target image. These models are often used to transfer the styles of famous artists to today's paintings. However, GAN modules work with very large data sets. This can cause the training time of the model to increase too much. All pictures of these artists are usually inserted into the model as a train set. In this study, we discussed the possibility of having more than one style in the paintings of the artist to be transferred in style and this situation may affect the model. We examined the effects of this situation on the cycle-GAN model. In our study, we compared the models trained with the whole data set, with the data entering this model trained with clusters after clustering. We used the K-means model and feature extraction methods to cluster the data set. We observed how clustered data affect the success of generative art models, and we aimed to reduce the training times of the models since we use smaller data sets.

1. Introduction

A lot of work has been done in the field of computer vision so far. Image to image translation is one of the core tasks in that area in a way that one of the source images is translated to the target image while keeping the originality of the source image. For this specific purpose of task, Generative Adversarial Network (GAN) is a helpful idea for

image to image translation. These networks are actually a combination of two networks that are Generator and Discriminator. We are using CycleGAN (Zhu, Park, Isola, and Efros, 2017) that is a technique for training unsupervised image translation by using GAN architecture for unpaired image-to-image translation. A CycleGAN is composed of 2 GAN's. That means a CycleGAN has 2 generators and 2 discriminators in total. One of the generators takes the images as an input from first dataset, and outputs images for the second dataset. After that, the other generator takes images from the second domain as input and generates images for the purpose of the first domain. Discriminator models are used to determine how appropriately generated images are created and update generator models according to these determinations. In this project, we are also using an additional extension of the CycleGAN that is called cycle consistency loss. This is essentially based on the purpose that the image output of the first generator can be used as the input of the second generator and that the output of this generator also matches the original image. For this purpose, we have calculated the cycle consistency in order to find the differences between real photos as input and transformation of generated Van Gogh images by using the input. Then, we have used the loss values in the calculation of the gradient.

We trained the cycle-Gan model with clustered datasets instead of skewing with the whole dataset and observed the results. Here are the expected results from observations; When we cluster structural as structural, it is better to transform a test picture of that struct; If an artist has more than one style, we can aggregate them based on styles to get better output from GAN and to reduce the training time of the model as we shrink the datasets. In order to do these clustering operations, we had to extract features from the train set we have. We used VGG19 feature extraction methods to do this. VGG is basically a convolutional neural network model. The numbers 19 represent how many convolutional layers there are in the model. This structure is generally used for image object classification(Rashid, Khan, Alhaisoni, Wang, Naqvi, Rehman, and Saba, 2020). In our study, we tried to use these VGG layers for both object-based classification and style-based classification. We have run the K-means clustering method on the features we obtained from feature extraction. With the clusters we obtained, we trained our GAN models and observed the results. K-means is basically

Tune It Up: Music Genre Transfer and Prediction

Fidan Samet¹ Oguz Bakir¹ Adnan Fidan¹

Abstract

Deep generative models have been used in style transfer tasks for images. In this study, we adapt and improve CycleGAN model to perform music style transfer on Jazz and Classic genres. By doing so, we aim to easily generate new songs, cover music to different music genres and reduce the arrangements needed in those processes. We train and use music genre classifier to assess the performance of the transfer models. To that end, we obtain 87.7% accuracy with Multi-layer Perceptron algorithm. To improve our style transfer baseline, we add auxiliary discriminators and triplet loss to our model. According to our experiments, we obtain the best accuracies as 69.4% in Jazz to Classic task and 39.3% in Classic to Jazz task with our developed genre classifier. We also run a subjective experiment and results of it show that the overall performance of our transfer model is good and it manages to conserve melody of inputs on the transferred outputs.

1. Introduction

"Music is the food of soul," said Arthur Schopenhauer. Music is an art that appeals to everyone. People have particular interest in music of certain genres such as Jazz, Pop, and Classic. By performing music genre transfer, music from different genres can be transferred to a certain genre. Thus, new hit songs can be generated automatically, musicians can easily cover music of different genres, and the music arrangements required in this process can be reduced. By music genre prediction, the performance of music genre transfer methods can be assessed. Furthermore, recommendations can be made to the listeners based on their favorite music genres. Thus, companies can obtain more profit with useful and accurate recommendations.

Style transfer between different domains has become a hot topic in the machine learning research field. Many deep

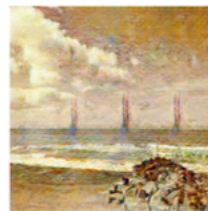
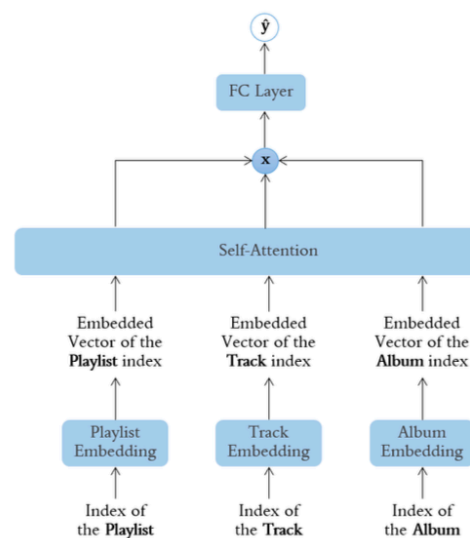
¹Department of Computer Engineering, University of Hacettepe, Ankara, Turkey. Correspondence to: Fidan Samet <fidansamet@gmail.com>, Oguz Bakir <oguzbakir0@gmail.com>, Adnan Fidan <adnfm@gmail.com>.

learning methods have been developed to accomplish this task. By extracting fundamental knowledge about domains with obtaining deep comprehensions, deep learning models can perform style transfer. Therefore deep generative models like Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) perform well in style transfer. Pioneering works have been done especially in the image domain, so that transforming images of summer to winter, night to day and photos to certain painter's drawings can be done successfully. In this project, we focus on style transfer in the music domain.

There are several music datasets in the literature. However, lyrics and various instruments are mixed in some of them. To perform music genre transfer, these variants must be separated from each other. Therefore, due to time issues, we consider a music dataset containing only piano as an instrument. We work on transferring these symbolic music from source to target music genre domain. Hence, we use one of the state-of-the-art deep learning methods, CycleGAN (Zhu et al., 2017) which performs unpaired image-to-image translation using cycle consistent adversarial networks, as our baseline. After adapting CycleGAN framework to music domain, we improve our baseline by adding auxiliary discriminators and triplet loss. Thus, we perform music genre transfer so that melody of the source genre retains while note pitches change according to the target genre. To assess the performance of the model, we perform genre prediction on the transferred music and check the classification accuracy according to the target domain. To that end, we test several machine learning classification algorithms and choose Multi-Layer Perceptron classifier, which gives the best accuracy. Since it is a challenging task to evaluate transfer methods, we also run a subjective experiment.

2. Related Work

(Brunner et al., 2018) introduces a method to perform music genre transfer on three major musical styles which are Jazz, Classical and Pop, by adapting CycleGAN model. They create and use a dataset consisting of MIDI (Musical Instrument Digital Interface) files. They introduce additional discriminators and classifiers to their CycleGAN-adapted approach. They train separate classifiers and transfer models for music genres. Their overall performance in both tasks is



Sample projects from Fall 2021 (AIN311)

— Multimodal ML —

Multimodal Image Retrieval

Berke Bayraktar¹ Yusuf Seven¹

Abstract

The task of image retrieval concerns many people on a daily basis. Whenever we would like to search for an image we interact with an image retrieval system. Initial systems leveraged tag based searches for retrieving images from a database [1, 2]. Later with image based methods, where features from the image were used for retrieval, were studied [3, 4]. However a more intuitive and effective way to search for images (especially products) are still an active research area; namely, using multiple modalities as a search query when searching for a target image. In this work we will study multimodal image retrieval where users can input an image and text describing modification on this image as query to obtain a target image.

1. Introduction

Ever since search engines have become popular there has been a need to provide a way for users to search for images just like they could search for text and webpages. One of the earlier approaches were tag based systems [1, 2] where images in the dataset would be manually tagged and whenever a user provided text query as an input, the system would compare this text to tags from the database. Problems with this approach is that information that can be obtained from the image itself is completely disregarded, moreover considering abundance of image data on internet, manual tagging is not a feasible approach. With the increase of image data and advances in computer vision, image based search methods were proposed [3, 4] where the query for a target image would also be a target image. Such retrieval approaches are widely used today. However a more intuitive and effective way to search images is to make use of multiple modalities and combine the use of text and image when generating a query for the target image as can be seen on figure (x). Leveraging multiple modalities has already been proven to be quite effective on variety of other tasks such as audio-visual speech recognition [5] video summerization [6] audio-visual emotion recognition [7] and many more, hence this research area has recieved a lot of attention over years. We would like to consider multimodal learning approaches for the problem at hand, namely image retrieval. To more formally

Figure 1.2: attribute modification retrieval samples

define this problem, we would like to retrieve a target image x_t given some query q_i where $q_i = f_{combine}(x_i, t_i)$ where x_i is a source image provided by user and t_i is a text describing modification on this source image, again provided by the user. In our work we are more focused on learning the composition function $f_{combine}$, the problem of retrieving x_t given q_i is a separate problem but in our case we will use metric learning approach. Details of this process will be discussed in section 3. We hope this work could be an introductory resource for those who are starting with multimodal machine learning and also showcase how leveraging multiple modalities could increase performance compared to using only a single modality for the case of image retrieval task.

2. Related Work

The main part of our work which is to obtain a multimodal representation for the features x_i and t_i has been studied extensively both in the context of image retrieval [8, 9, 10] and for other tasks as well such as visual question answering [11]. However we will specifically consider image retrieval works. In [8] Han et al. propose a automatic spatially-aware concept discovery approach. And use these concepts for product retrieval on Fashion200k dataset which they introduce themselves in same work. This dataset contains 200k images of fashion products of 6 classes: dresses, jackets, pants, skirts and tops. Dataset also contains captions that explain these products. The dataset will be explained more detailly in section 4. In this work authors first train

Audio Calptioning

Alperen Ozcelik¹ Yunus Alper Bagcilar¹

Abstract

Audio captioning is focusing on identifying the human-perceived information in a general audio signal and expressing it through text, using natural language. This information includes identification of sound events, acoustic scenes, spatiotemporal relationships of sources, foreground versus background discrimination, concepts, and physical properties of objects and environment. In this project, we aimed to design a CNN-based model for audio files feature extraction and a transformer-based NLP model for caption generation. We evaluate our model on Clotho dataset's test set, which has approximately 1,000 audio files, along with 5 captions for each audio file.

1. Introduction

For us, audio signals are one of the most important sources to receive information from the outside world in our daily life. Automated audio captioning is a multimodal problem in which the model takes audio signals as inputs and gives a natural language description of that audio signal. Understanding only the acoustic events isn't enough to solve the audio captioning problem. The audio captioning models also have to understand natural language and some other pieces of information from the audio signal such as discrimination of foreground from the background (e.g. Cars driving by outside with heavy rain falling), space-time relationships of audio sources (e.g. A car with a siren passes by and then a bicycle passes by), and physical information of objects and environment (e.g. Droplets of water are falling on a metallic surface). Therefore, audio captioning can be useful in a variety of areas, such as helping the hearing impaired understand sounds and sound-based security systems.

2. Related Work

In the past studies, deep learning approaches based on encoder-decoder architecture have been presented for the automated audio captioning problem. The first approach to Automated Audio Captioning was proposed by Drossos et al., (2017). Interest in automated audio captioning has grown with the release of two new freely available datasets Clotho(Drossos et al., 2020), AudioCaps(Kim et al., 2019).

Le Recommendeur: A Music Recommendation System Enhanced by NLP

Abdullah Palaz¹ Mete Mert Birdal¹ Gökalp Özer¹

Abstract

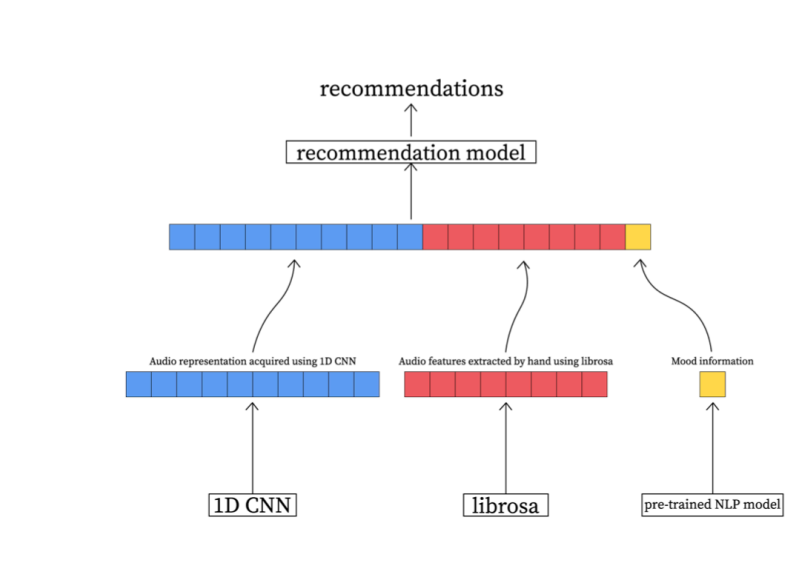
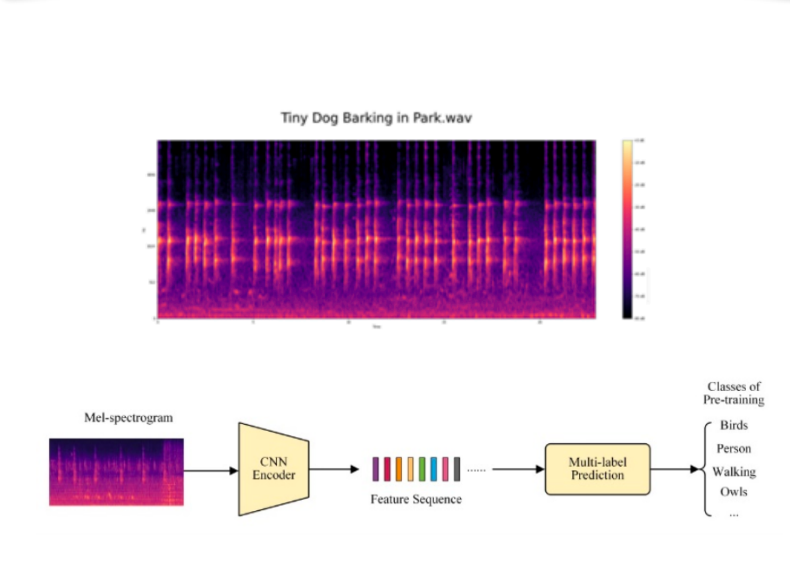
The vast amount of available data creates the problem of data overloading. In this environment, finding relevant/important/needed data points (i.e. websites, products, videos, or songs as in our case) is hard and is a problem that needs to be tackled. It is being tackled using recommendation systems. Music is an interesting and needed domain to be worked on because we listen to music all day and every day. Not just every day but everywhere, doing everything too. We work and listen to music, we are at the gym, working out, we listen to music, we study and we listen to music, we travel and we listen to music and this list can go way longer. But listening to that much music creates a problem where this particular person that listens to too much music becomes new musicless. This is the aforementioned problem projected in the music domain. So, it is an urge to find new music to listen to. This project's goal is to recommend you music that you would like. As known, music is just audio-and, audio is feature-rich data. Our approach is content-based. We extract audio features from songs' audio and lyrics to recommend songs. This paper structured as follows: Section 2 discusses work on recommendation systems and music recommendation systems. Section 3 introduces our dataset, Section 4 explains our approach, Section 5 shows and discusses our results and Section 6 summarizes our work.

1. Introduction

The amount of data creation increased vastly, with the digitalization of everything. As much as 64.2 Zettabytes of data is estimated to be created in 2020 and 181 Zettabytes of data creation is estimated for 2025 (1 Zettabyte = 1e+12 Gigabyte). Even though only 2% of the data created in 2020 was stored and retained into 2021 (Holst, 2021), and this vast amount of available data creates the problem of data overloading. In this environment, finding relevant/important/needed data points (i.e. websites, products, videos, or songs as in our case) is hard and is a problem that needs to be tackled. This problem is being tackled using recommendation systems. These recommendation systems

2. Related Work

Since recommendation systems are being used in various domains, there are a few approaches to them: collaborative filtering, content-based, utility-based, demographic-based, knowledge-based and hybrid-based (Fayyaz et al., 2020). In the music domain, collaborative filtering and content-based approaches are widely used as well as hybrid models. Recently, researches on this domain try to include multimodalities mostly using emotion extracted from different sources. (Chang et al., 2017) try to recommend songs for stress relief and to do that, use electroencephalography feedback as an enhancement to the classical methods. In 2018 (Ayata et al., 2018) enhanced music recommendation algorithms by using the "mood" of a person (extracted by Wearable Physiological Sensors). In 2019, (Chang et al., 2017) extracts emotion from the music's audio itself. In



Sample projects from Fall 2022 (AIN311)

— ML for Education —

Project LEAFS: Learning Efficiency Assessment from Footage of Students

Abdullah Enes Ergun¹

Baha Kirbasoglu¹

Can Ali Ates¹

Abstract

Attitudes of the students in class affect lecture efficiency for both the lecturer and themselves. Keeping track of each student's attitude simultaneously during to lecture is a really hard problem. Therefore, we propose a method to detect different attitudes of students for solving this problem in this case study. We use the YOLOv5 tool that combines deep learning and computer vision techniques. For instance, Convolutional Neural Networks. We collected our dataset using web scraping, taking classroom photos by hand and using a small portion of dataset which is available online. The dataset which is collected from the multiple sources contains seven classes. These classes are separated into three positive and four negative classes. We created a semi-theoretical formula based on these classes that calculates lecture efficiency according to detected class instance counts and weights of these classes. While positive classes increase the efficiency during lecture efficiency assessment, on the other hand, negative labels decrease. Our goal is to help lecturers to improve their lectures, based on students' attitudes.

1. Introduction

Most of the time, assessing the efficiency of a lecture for students is a problem for the lecturer. The lecturer cannot be sure if the lecture is understandable or not. In this situation, the lecturer needs a simultaneous system to keep track of the students one by one based on their attitudes. Students can act different behaviours during the lecture such as playing phone, taking notes, sleeping, etc. These attitudes can have different meanings. For instance, a student who yawns is not completely distracted from lecture, can still listen the lecture without hundred percent attention. The lecturer cannot assess all of these attitudes in real-time during a lecture, so an automated system has to detect students' attitudes and evaluates the status of the student based on what the student does. Therefore, developing a system that calculates the lecture efficiency based on semi-theoretical formula that created by ourselves with giving weights to these student attitudes and then reports the calculated lecture efficiency

with detected student attitudes in class both visually and written to the lecturer in real time may help both the students and the lecturer. This developed system is supported with graphical user interface, in other words GUI, that is designed by us to create user friendly environment. Lecture efficiency basically can be assessed with seven attitudes. These attitudes can be evaluated as positive and negative. While positive attitudes such as listening, taking notes and raising a hand increase lecture efficiency; on the other hand, negative attitudes such as yawning, playing with the phone, sleeping and eating or drinking something can decrease lecture efficiency. Weights of these attitudes are determined by us scaled into zero-one range. The algorithm that we use assesses the lecture efficiency and returns a basic report to a lecturer to prevent inefficient lectures. As a result of these reports, the lecturers can change their own teaching techniques or materials, and students can focus more with these improvements then change their own behaviours.

2. Related Work

Assessing student attitudes is a rare research topic that we found a few articles about. There are several articles use different techniques to detect behaviour of students. Article¹ collects their own data from classroom videos like we made. After that, combines the temporal and action detection to create recognition model. Then, uses this recognition model for task recognition. It uses the recognition results by giving these results to different type of video captioning models such as, HACA and RecNet. This study had some troubles like ours such as misclassifications and lack of dataset. For instance, model detects reading book as playing with phone or vice versa. So, some of the misclassification occurs because of the perspective. This study provides us a perspective to detect student behaviours and misclassifications in different way, so this article can use for the future directions of our model. Article^{2,3,4} is about detection systems to detect mobile phone usage of a person with using classical Convolutional Neural Network and Faster Region Based Convolutional Neural Network. Article⁵ uses mAP@0.5 as the evaluation metric. This studies, give us an approach about detection of behaviours which can be identified by pairing another tool such as taking note detection with pencil and notebook, playing phone detection with phone, eating or drinking detection with foods and beverages. Also, ex-

Question Assistant Barlas

Mehmet Berat ERSARI Zeynep Hafsa Dilmaç

Abstract

We need machine learning techniques to automatically generate questions from educational materials such as textbooks and articles. This is an important problem in education because it can help students better understand and retain the material they are learning. In order to generate questions, an education AI system must be able to understand the underlying concepts and ideas presented in the material, as well as the appropriate level of difficulty for the questions. Research in this area is ongoing and aims to develop algorithms and models that can generate high-quality questions that are both challenging and relevant to the material being studied. In this project, we realize question generation, which we think will be very useful in the field of education. We use mT5 (Xue et al., 2021) models while executing this artificial intelligence problem. While approaching this problem, our passion and the main goal is to produce Turkish questions from Turkish texts. We find TQuad (TEKNOFEST, 2018) dataset, which we used and fine-tuned, from Teknofest's website before. We get the listed scores after fine-tuning on mT5 small and base model. The best results we have got in mT5-base for 20 epochs; BLEU-1 0.24, BLEU-2 0.31, BLEU-3 0.38, BLEU-4 0.42, ROUGE-1 0.43, ROUGE-L 0.44, METEOR 0.32.

1. Introduction

We are embarking on a project to create a question-generation model, which is a basic problem in the field of artificial intelligence. This project was interesting because in our opinion, one of the most important areas of artificial intelligence is NLP. Artificial Intelligence needs to be able to understand and interpret human language. This requires deep learning models. In order to train this model to accurately generate questions, we need a large amount of labeled data in the form of paragraphs with related questions. Our goal is to develop a system that can predict a relevant question based on a given paragraph in Turkish. While there is a wealth of training data available in English, such as the WebQuestions dataset (Talmor & Berant, 2018). Most of the research in this area has focused on the English language.

2. Related Work

When text generation is researched as a topic, it is seen that there are different subjects in this field and many studies on these subjects. In a study conducted in 2022, a text generation study was carried out to increase text classification (Bayer et al., 2022). In another recent study conducted in 2022, a text summary study was conducted for Italian texts and good results were obtained with BART-IT model (La Quatra & Cagliero, 2022) Many more such examples can be found. If we narrow the subject a bit, a lot of work has been done on question generation and question answering. In the past, it has focused more on answering questions and therefore question answering studies are more than question generation studies. Question generation has been a somewhat more up-to-date subject of study. Question answering, on the other hand, still has current studies. In the study conducted in 2022, question answers were created from Dutch texts using the BERT model (de Jong & Bouma, 2022). Finally, when the question generation studies, which is our subject, are examined, it is seen that in a study conducted in 2020, question generation was made from the texts, the

Career Path Predictor

Melike Nur Dulkadir, Sare Naz Ersoy

Abstract

Choosing a career path in computer science is an important decision that can have a significant impact on every individual future. A career in computer science offers many exciting and rewarding opportunities, as well as the potential for high salaries and job security. However, it is also a field that is constantly evolving and requires ongoing learning and development. There are many different fields and specialties within computer science. Each of these fields requires a different set of skills and knowledge, and it is important to choose a path that aligns with your interests and strengths. With this in mind, we have developed a career path recommendation system as a solution to this problem. This program utilizes various machine learning techniques, including Gradient Boosting, AdaBoost, and Neural Network, to recommend career paths by combining two datasets. The larger of these datasets currently contains 20000 rows and 39 columns. Our aim is to provide helpful career path recommendations to users.

1. Introduction

It has become common for students to select their career paths based on the choices of their friends or the friend that offer the highest salaries rather than considering their strengths. This approach often leads to disappointment and disillusionment. Additionally, when hiring candidates, recruiters must evaluate them on various factors. Therefore, there is a need for a system that can assist students in choosing a job role that is well-suited to their skillset and other evaluation criteria, which can now be achieved through the use of machine learning techniques. A career path recommendation system is a tool that uses data and algorithms to suggest potential career paths based on an individual's interests, skills, and experience. These systems can be incredibly useful for individuals who are unsure of what they want to do with their careers, who are looking to make a change but are not sure where to start. By analyzing data on job demand, salary, and other factors, a career path recommendation system can provide tailored recommendations that can help

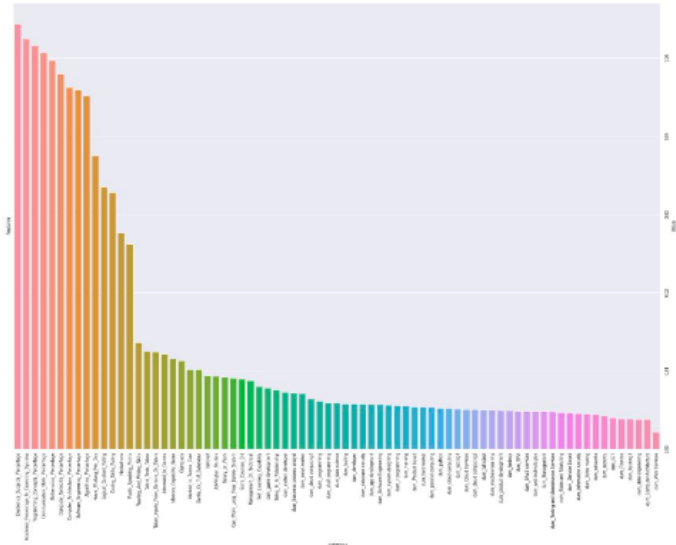
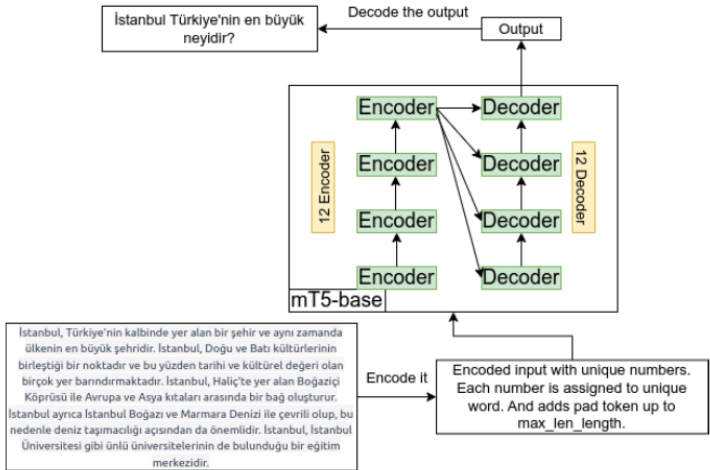
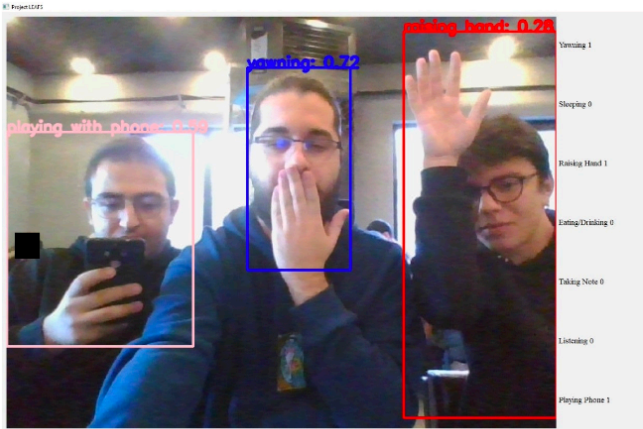
2. Related Works

The selection of a career path is a crucial decision that can have a significant impact on an individual's future. Therefore, numerous studies have been conducted on this topic. In this review, we examined four studies that focused on predicting future career paths using machine learning techniques.

The first of the articles containing studies similar to our project is Skill-based Career Path Modeling and Recommendation, created by Aritra Ghosh, Beverly Woolf, Shlomo Zilberstein, and Andrew Lan (1) . In this article, the Monotonic Nonlinear State-space Model was created using real-world datasets (LinkedIn and Indeed) to analyze online user professional profiles and provide users with actionable feedback and recommendations on how to achieve their career goals. This model offers users a path that includes intermediate steps to achieve the purpose they specify.

The second article is A Machine Learning Approach for Future Career Planning, created by Stanford University students Yu Lou, Ran Ren, Yiyang Zhao (2) . In this study, each person's profile and career path are considered as a sequence containing more than one node, and each node is represented by various polynomial features (vi1, vi2, ..., vik). Considering a person's current career path and goal, it is aimed to propose the most appropriate career path, that is, the path with the highest probability of reaching the target node, using the Markov Chain model. In addition, the K-means clustering algorithm is used to group many semantically similar location headers.

The third article, which we consider as the primary example for our study, is the article titled Computer Science Career Recommendation System Using Artificial Neural Network by Brjmhohan Daga, Juh Checker, Anne Rajan, Sayali Deo (3). The dataset used in this study is the same as the dataset of our main study that we used when presenting our first proposal. In this article, using 20000 observations, academic percentages, and extra & co-curricular activities, various vocational suggestions are made for people according to their various characteristics such as activities and personal choices. An artificial neuron network (ANN) model was



Sample projects from Fall 2023 (AIN311)

— ML for Sustainability —

Predicting the Clearsky GHI for solar panels

Kemal Şahin¹ Tuncer Sivri² Burak Kurt³

Abstract

This research focused on predicting Global Horizontal Irradiation (GHI) which is a crucial factor in solar energy systems optimizations. Various machine learning models and algorithms including Linear Regression, Random Forest, AdaBoost, 1D-CNN, and LSTM, were investigated and experimented with. The study revealed that 1D-CNN with optimized parameters outperformed other models, achieving a minimal MSE of 3.9×10^{-5} . However, the Weighted Voting model combining different models did not yield superior results compared to 1D-CNN alone. The findings suggest that in order to predict GHI in solar energy applications 1D-CNN stands out as a reliable and effective model superior to others.

1. Introduction

The total amount of solar radiation that reaches the Earth's surface per unit area and time is called Global Horizontal Irradiation (GHI). It is a factor that affects the design and operation of solar panels and other renewable energy systems, as the potential power output and efficiency of these devices are determined by it. The prediction of GHI is influenced by many complex and dynamic factors, such as cloud cover, aerosols, humidity, and atmospheric conditions. Therefore, accurate and reliable GHI forecasting methods are needed to optimize the planning and management of solar energy systems. Machine learning techniques, which are methods that learn from data and make predictions based on patterns and relationships, are one way to approach GHI prediction. Nonlinear and high-dimensional data can be handled by machine learning, and it can adapt to changing environments and scenarios. Different machine learning techniques for GHI prediction, using weather data as input features, will be explored in this report. The performance of linear, nonlinear, and sequential models, such as linear regression, Random Forest, and recurrent neural networks, will be compared. The advantages and limitations of each technique will also be discussed, and possible improvements and future directions will be suggested.

2. Related Work

The problem of optimizing renewable energy sources for sustainable development has been addressed by various researchers using different methods and perspectives. In this section, some of the related works will be reviewed, and highlighted their contributions and limitations. Researches show that some machine learning methods that are not based on neural networks can perform well on this task. In (Feng et al., 2019), researchers used a random forest algorithm along with neural networks. Results are very close, even the RF algorithm outperformed NN-based models in some cases. K-nearest-neighbor algorithm also performs well on image data with proper data preprocessing (Pedro & Coimbra, 2015). However, most of the time neural networks are preferred to solve this problem. In (Chiteka & Enweremadu, 2016), researchers developed a very simple neural network with one hidden layer and seven input layers. Its performance is the best of all the models stated but its limitation is that model only works well on small feature space. For more complex inputs like images, CNN-based models are proposed in (Yang et al., 2021). Researchers developed a 3D CNN model to obtain features from image data and they used the obtained features to train their SVM, ANN, and KNN models. Results show that the CNN structure is suitable for feature-extracting tasks in this domain. Lastly, to be able to capture both spatial and temporal features, the CNN-LSTM model is proposed in (Zang et al., 2020). The proposed model first applies a convolutional neural network (CNN) to extract spatial features from a two-dimensional matrix composed of meteorological parameters associated with a target site and its neighboring sites. Then, a long short-term memory (LSTM) network is applied to extract temporal features from historical solar irradiance time series data associated with the target site. Each model explained here has strengths and weaknesses. In this study, we aim to use these advantages and drawbacks and develop a model that describes the given data best.

3. Methodology

3.1. Mean Square Error (MSE)

Mean Square Error (MSE) is a performance metric commonly used in predictive modeling to assess the accuracy of a model's predictions. MSE is used to evaluate our model's

1

Predicting Real-Life CO₂ Emissions of Passenger Cars: A Machine Learning Approach

Bengü Barış Balkan^{*} Furkan Özyurt^{*}

Abstract

As the climate change becoming a global problem over the years, reducing CO₂ emission became a primal goal. Passenger cars are a significant contributor to CO₂ emissions, making it essential to understand and predict real-world emissions accurately. This project aims to predict real-life CO₂ emissions of passenger cars using given laboratory test information and real-world user-submitted or automatically recorded data.

1. Introduction

CO₂ emissions of the cars are measured as part of the vehicle certification [1] test which is Worldwide harmonized Light vehicles Test Procedure (WLTP). Although this test is widely used, there is increasing evidence to suggest that officially reported CO₂ values do not reflect the actual performance of the vehicles on the road. There are reports that shows 12%-44% [2,3] divergence between official values and real-world estimates. Reducing this divergence can provide better regulations on car manufacturers and benefit the environment.

This project aims to reduce this divergence by predicting on-road CO₂ emissions of the passenger cars using machine learning models. To consider most of the possibilities we've used linear, regularized linear and non-linear regression models.

The data-set used in the project consists of cars' mechanical attributes like axle-width, engine power, mass etc. It has 214,858 instances with 9 features and a real-world CO₂ emission value to predict.

Average root-mean-square errors of all models is 35.64. Mean of the real-world estimated CO₂ emission values is 168.72. Which yields 21.12% divergence between the predictions and real-world estimations. This result stays in the lower half of the reported divergence and can be considered moderately successful.

It is environmentally crucial to reduce CO₂ emissions of the passenger cars. Improving the certification process of the cars should be the main step of this goal. However,

until necessary regulatory steps are taken, machine learning approaches can be used to enhance the test results.

2. Related Work

The divergence between official CO₂ emission values and real-world estimates is known for years, [4] However, recent studies show that the divergence did not get smaller in past years. In the opposite direction, it has increased over time, from approximately 9% in 2001 to 42% in 2016 [5].

There were non-machine-learning attempts to predict the real-world emissions of passenger cars. [6]. Mock et al. used statistical models which used in econometry and showed the divergence of real-world emissions and manufacturers' type-approval CO₂ results in 2014.

More recent studies use more complex models such as neural networks and semi-supervised clustering to predict the divergence and solve other problems. Maino et al. [7] uses a deep neural network-based model for the prediction of hybrid electric vehicles (HEV) carbon dioxide emissions. In their research they have average regression errors below 1%. Their presented model also aims supporting HEV design optimization phases.

Niroomand et al. use novel semi-supervised fuzzy C-means (SSFCEM), random forest and AdaBoost models to classify passenger vehicles and enable segment-based CO₂ emission evaluations.[8] Their research applies deep learning techniques to evaluate inter-class and intra-class vehicles differences in CO₂ emissions.

Also, there are studies which does not try to predict the divergence between real-world emissions and test results. For example, Mock et al. suggest new test procedures to improve accuracy in the laboratory results [9]. They mention differences between differences between WLTP and NEDC which are new and old test procedures.

1

Air Quality Prediction With Machine Learning Algorithms

İzzet Ahmet , İlbey Gülmez

Abstract

In the existing literature on air quality, there are numerous studies focused on developing new formulas for calculating Air Quality Index (AQI). The study tries to help calculate AQI at a lower cost by focusing on estimating CO values related to air quality without the need for measurement tools and at minimum cost. The primary aim is to quickly obtain CO data using machine learning methods such as Support Vector Regression, KNN Regression, Multiple Linear Regression, DTR and RFR rather than calculating the Air Quality Index (AQI). The study compares the effectiveness of reference analyzers (GT) and nominally targeted (PT08) meters. The success rate of the study was reported as 69% and the mean square error (MSE) was reported as 29%.

1. Introduction

Air pollution has emerged as a worldwide concern, causing diverse health and environmental challenges. The rise in industrialization, agricultural practices, urban development, and heightened fuel consumption driven by recent technological progress has resulted in environmental issues like water, noise, and air pollution. Air pollution, in particular, directly affects human health through pollutants and particulate matter. Consequently, there is a growing interest within the scientific community to research air pollution and its consequential effects. [1].

Researchers have recognized that historical monuments face potential deterioration due to the adverse effects of air pollution [2]. The presence of greenhouse gases has a detrimental impact on climate conditions, influencing plant growth negatively [3]. Emissions of inorganic carbons and greenhouse gases further disrupt plant-soil interactions [4]. Climatic fluctuations not only impact human health and animal life but also exert significant influence on agricultural aspects and productivity [5]. This situation leads to economic losses, with estimates from the American Lung Association indicating an annual cost of approximately 37 billion dollars in the US related to illnesses associated with air pollution.

California alone accounts for 15 billion dollars of this total [6].

Conventional approaches relying on probability and statistics are intricate and comparatively less efficient. The effectiveness of Machine Learning models in predicting the Air Quality Index (AQI) has been demonstrated to be more dependable and consistent. The advent of advanced technologies and sensors has simplified and enhanced the precision of data collection. Achieving accurate and reliable predictions with extensive environmental data demands thorough analysis, a task efficiently managed by ML algorithms [7].

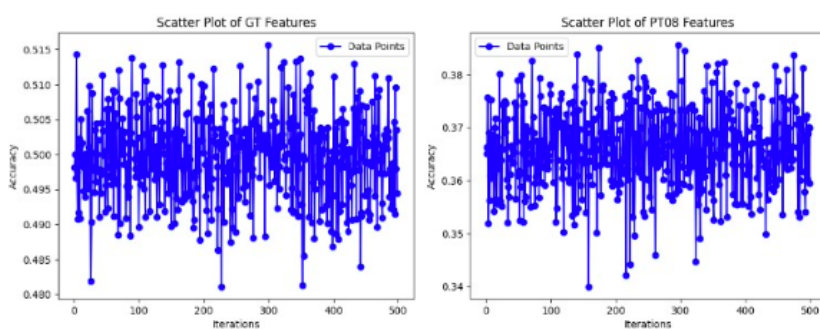
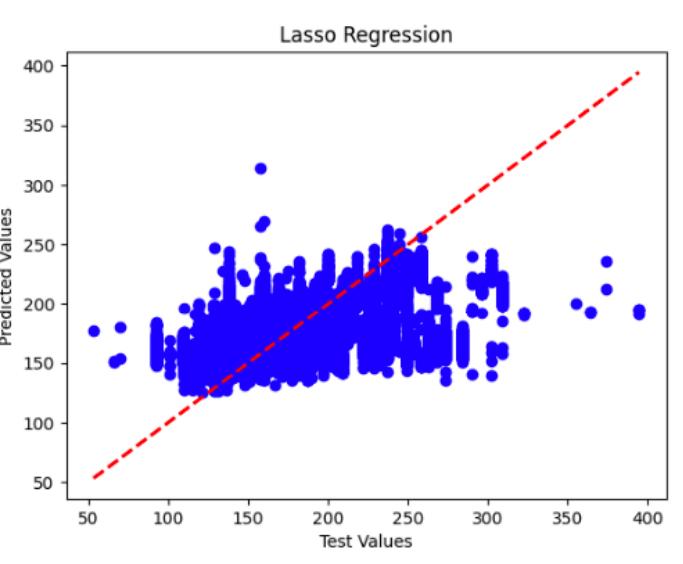
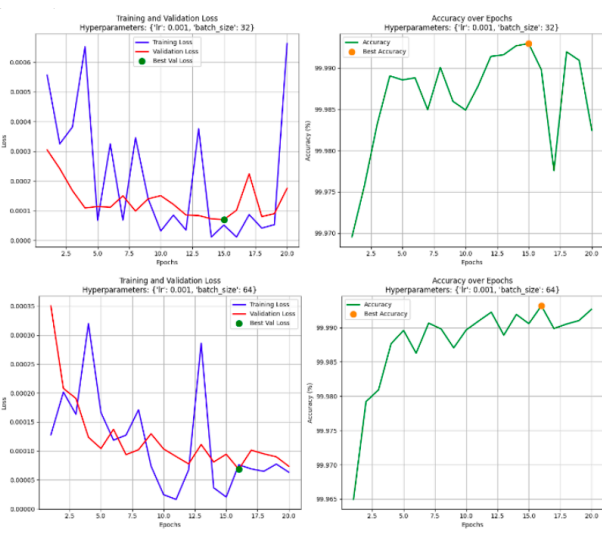
This research aimed to develop predictive models for the concentration of CO in the air, utilizing the Air Quality Dataset specific to the city of Cassino in Italy. Various commonly employed regression analysis techniques, such as KNN Regression, Support Vector Regression (SVR), and Multilinear Regression, were utilized. Our approach involved constructing hourly models for each measurement related to pollutants and particulates, in addition to a model predicting the hourly Air Quality Index (AQI) for Cassino. The primary objective of these models was to forecast air pollution levels in the region with minimal reliance on measurement instruments, ensuring more cost-effective and accurate predictions. This facilitates the anticipation of potential adverse effects on community health and the timely detection of such impacts, duly reported.

The second section of the study provides an analysis of machine learning methods employed in the literature for air quality prediction, accompanied by a literature review. Section 3 details the dataset under examination, outlining preprocessing steps and feature selection techniques. Section 4 includes a comparison of data resulting from experiments conducted with different models and parameters. Finally, in the concluding section, the obtained results are thoroughly discussed, leading to conclusive remarks.

2. Related Works

Various methodologies have been explored in the literature about the Air Quality Index (AQI), and these can be broadly classified into three main categories. The first involves the analysis of sensor data through techniques like time

1



Sample projects from Fall 2024 (AIN311)

— ML in Sports —

Predicting NBA Player Injuries: A Data Driven Approach

Süleyman Yolcu^{*1} Emirhan Utku^{*1}

Abstract

Injuries are a pervasive concern in professional sports, with the National Basketball Association (NBA) offering a unique context for examining this issue due to its high intensity and physical demands. This project aims to develop a predictive model for NBA player injuries using machine learning techniques. By integrating historical injury data with player performance metrics, we strive to identify patterns and factors that contribute to injury risk. Our preliminary findings suggest that decision tree models show promise in accurately classifying injury-prone players, thereby paving the way for improved injury prevention strategies, enhanced player longevity, and more informed team decision-making.

1. Introduction

Injuries in the National Basketball Association (NBA) significantly impact players, teams, and the league as a whole. From a player's perspective, injuries can derail promising careers and shorten playing longevity. For teams, unanticipated injuries to key athletes can drastically alter season trajectories and championship aspirations. At the league level, high-profile injuries can affect viewership, revenue, and the overall quality of competition.

Over the past decade, the growing sophistication of sports analytics has ushered in new approaches to understanding and mitigating injury risk. Machine learning (ML) and data-driven techniques now enable the integration of large, heterogeneous datasets—including game statistics, biometric measurements, and training loads—to identify hidden correlations and patterns associated with injury occurrences. This data-centric methodology has far-reaching implications. By revealing potential risk factors, ML models can help coaches and medical staff optimize training regimens, manage player workloads, and implement targeted interventions tailored to each athlete's profile.

In the NBA's fast-paced environment, factors such as minutes played, back-to-back games, and accumulated fatigue can expose players to increased injury risk. Consequently,

developing robust predictive models that account for such situational stressors represents a pivotal step toward reducing downtime and extending playing careers. Ultimately, the capability to predict which players are more susceptible to injuries and when they might occur has the potential to revolutionize player management strategies, minimize disruptions to team rosters, and sustain higher levels of competition for fans.

This project aims to develop a machine learning-based framework to predict NBA player injuries using historical injury records combined with in-game performance metrics. By identifying patterns and correlates of injury risk—such as workload, minutes played, player physical attributes, or specific types of injuries—coaches, trainers, and team decision-makers could proactively manage player rest, recovery protocols, and conditioning programs. Ultimately, the goal is to enhance player well-being, guide strategic rotational decisions, and improve fan engagement by maintaining healthier rosters throughout the season.

2. Related Work

The field of sports injury prediction has witnessed significant advances in recent years, largely due to expanded data availability and sophisticated machine learning algorithms. Within the NBA context, a notable contribution comes from Farghaly and Deshpande (Farghaly & Deshpande, 2024), who applied models such as Random Forest, Gradient Boosting, and K-Nearest Neighbors to anticipate injuries. Their dataset emphasized physical measurements (e.g., player height, weight, and agility tests) as well as historical injury records. Notably, their Random Forest classifier attained an accuracy of 92.47%, illustrating the strength of ensemble methods in handling complex interactions among anthropometric data and prior injury history.

Although our approach shares similarities with Farghaly and Deshpande's work—particularly in combining historical injury data with player-specific attributes—it departs from theirs by incorporating a broader range of in-game performance variables, such as minutes played, pace of play, and distance covered. By capturing game-by-game trends, our study aims to identify acute stressors (e.g., back-to-back

SportVision

Emir Çağrı Kökdener^{*1} Abdullah Edik^{*1}

Abstract

This study aimed to classify sports images and automate the labeling process for real-world applications. We utilized Kaggle's sports dataset, which consists of 100 images across various categories. To achieve this goal, several approaches were explored, including ResNet trained from scratch, ResNet with transfer learning, Multi-Layer Perceptron (MLP), k-Nearest Neighbors (KNN), KNN with feature extraction using a ResNet50 model, Random Forest with feature extraction using a ResNet50 model, and a Convolutional Neural Network (CNN).

Among these methods, ResNet with transfer learning achieved the highest performance, reaching an accuracy of 95.00%. This result underscores the effectiveness of pre-trained models in tackling image classification tasks, particularly when working with limited datasets. The findings highlight the potential of transfer learning as a powerful tool for automating complex image labeling processes in real-world scenarios.

1. Introduction

As we stride confidently into the future, the integration of artificial intelligence has significantly accelerated our progress. The inclusion of AI has made data-one of the most valuable resources—even more indispensable. Data, which serves as the key to solving countless problems, is generated constantly as we go about our daily lives, contributing to a massive and ever-growing repository. While we may not always capture or organize this vast data effectively, what we do manage to store can be structured, processed, and analyzed to yield meaningful insights and outcomes.

In our project, we focus precisely on this potential, working with a dataset of sports-related images. Our primary goal

^{*}Equal contribution ¹Department of Artificial Intelligence Engineering, Hacettepe University, Çankaya/Ankara, Turkey. Correspondence to: Emir Çağrı Kökdener <emirkokdener@hacettepe.edu.tr>, Abdullah Edik <abdullahedik@hacettepe.edu.tr>.

is to automate the labeling of these images, as opposed to manually identifying each one. This research centers on the development of an efficient image classification system tailored to sports imagery. With the sheer volume of competitions and recordings across diverse sports, manually labeling these images is both labor-intensive and time-consuming. By leveraging machine learning techniques, our aim is to create a system capable of accurately and efficiently classifying sports images, drastically reducing the need for manual intervention and enabling faster data analysis. This will not only streamline the process but also unlock new possibilities for analyzing sports-related data at scale.

2. Related Work

Image classification is a crucial problem in computer vision, with significant research conducted in the area of sports image classification. Notable works include:

SE-RES-CNN Model: This model integrates a Squeeze-and-Excitation (SE) module to dynamically adjust feature map weights and enhance critical feature recognition. Combined with residual connections, it achieves high performance across multiple datasets without increasing complexity (Li et al., 2024).

Simultaneous Image Classification and Annotation: This study proposes a probabilistic model treating class labels as global descriptors and annotation terms as local descriptors, enabling simultaneous classification and detailed image annotation (Chong et al., 2009).

InceptionV3 and Neural Networks: A robust framework leveraging InceptionV3 for feature extraction and Neural Networks for classification, effectively handling diverse sports image categories (Joshi et al., 2020).

Bayesian Classifier: This method uses a Bayesian framework with four key color features to classify sports images, demonstrating the strength of probabilistic approaches (Jung et al., 2003).

These studies highlight diverse methods, from neural networks to probabilistic models, addressing the unique challenges of sports image classification.

Sports Video Classification Using CNN and RNN Models

Alperen Demirci¹ Bora Dere¹

Abstract

Classification is a fundamental task in machine learning, with applications in various domains (Wu et al., 2015). In this project, we explore the classification of sports activities from video data using a combination of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). We utilized a pretrained ResNet-50 model as the backbone for feature extraction from video frames. These features were subsequently fed into four different RNN architectures: Bidirectional LSTM, GRU, LSTM, and Residual LSTM, to capture temporal dependencies and improve classification performance. Among them, we have achieved the best result using the BiLSTM model with a test accuracy of 74.11%. Our findings highlight the effectiveness of integrating CNNs and RNNs for video-based sports classification, with notable performance differences between RNN architectures. Future work will focus on further optimizing the models and exploring additional techniques to enhance the accuracy of classification.

1. Introduction

Machine learning has revolutionized the way we approach complex problems, offering innovative solutions in various domains (Sarker, 2021). Among these, classification stands out as one of the most studied and applied tasks (Sooft & Awan, 2017), (Jijo & Abdulazeez, 2021). Achieving or even surpassing human accuracy in classification tasks has been a long-standing goal in the field, driving significant advancements in algorithms and techniques.

In this project, we focus on the problem of classifying sports types from video data. This is an intriguing challenge, because it requires understanding both the spatial characteristics of individual frames and the temporal dynamics across video sequences. The unique combination of these aspects makes video classification a compelling problem to explore.

To address this problem, we used the feature extraction capabilities of Vonvolutional Neural Networks (CNNs), which excel at capturing spatial features from images (Yue-Hei Ng

et al., 2015), (Varshni et al., 2019), (Jogin et al., 2018). Since videos are sequences of frames, we utilized a pre-trained CNN model to extract features from each frame. These frame-level features were then input into Recurrent Neural Networks (RNNs), which are designed to model temporal relationships in sequential data, enabling us to develop a robust video classification pipeline.

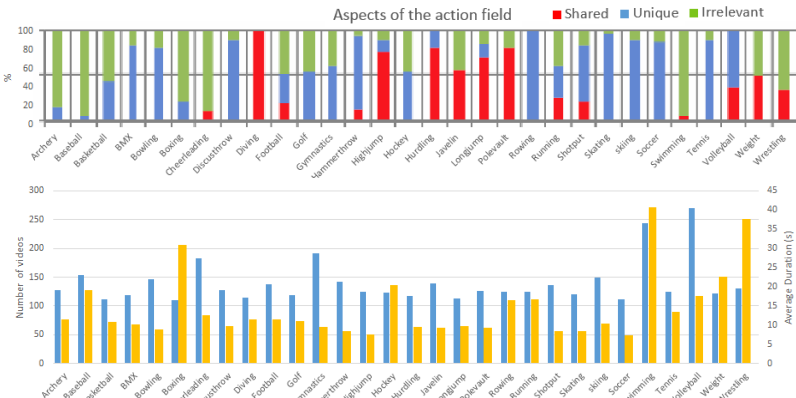
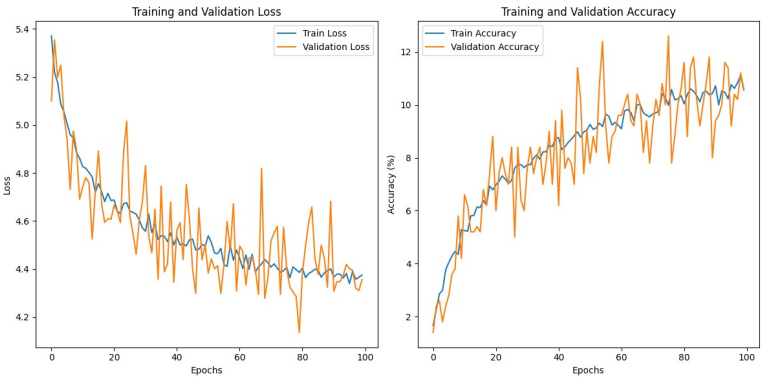
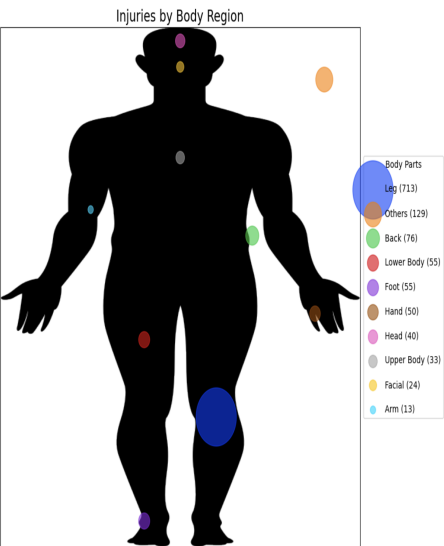
We find this problem particularly interesting because it allowed us to combine CNNs and RNNs, two powerful models we had previously studied in isolation, into a cohesive framework. This integration not only deepens our understanding of these architectures, but also demonstrates their complementary strengths in solving a complex problem like video-based sports classification.

2. Related Work

Video classification has emerged as a significant area of research, with various approaches proposed to address its challenges. One of the foundational methods involves the use of CNN only, focusing on fusion of information across the temporal domain (Karpthy et al., 2014). While this approach dates back to 2014 and might appear simplistic by today's standards, it introduced valuable concepts such as multi-resolution CNN architectures. These ideas laid the groundwork for more advanced temporal modeling techniques in video classification.

A closer approach to ours was proposed in 2015, which aimed to "harness not only the spatial and short-term motion features, but also the long-term temporal clues" (Wu et al., 2015). Unlike our methodology, this study employed a stacked hybrid architecture, utilizing CNNs to extract "two types of features...from spatial frames and short-term stacked motion optical flows respectively." This strategy effectively combined spatial and motion features, demonstrating the potential of hybrid architectures for video classification tasks.

Another study (Yue-Hei Ng et al., 2015) took a slightly different direction, using AlexNet and GoogLeNet models to process individual video frames. They extensively experimented with various pooling techniques, including Conv Pooling, Late Pooling, Slow Pooling, Local Pooling, Time-Domain Convolution, and GoogLeNet Conv Pooling.



Collaboration Policy

- All work on assignments have to be done **individually**. The course project, however, can be done **in pairs**.
- You are encouraged to discuss with your classmates about the given assignments, but these discussions should be carried out in an abstract way.
- **In short, turning in someone else's work, in whole or in part, as your own will be considered as a violation of academic integrity.**
- Please note that the former condition also holds for the material found on the web as everything on the web has been written by someone else.

<http://www.plagiarism.org/plagiarism-101/prevention/>

Course Outline

- **Week1** Overview of Machine Learning, Nearest Neighbor Classifier
- **Week2** Linear Regression, Least Squares
- **Week3** Machine Learning Methodology
_____ *Assg1 out*
- **Week4** Statistical Estimation: MLE, MAP, Naïve Bayes Classifier
- **Week5** Linear Classification Models: Logistic Regression,
Linear Discriminant Functions, Perceptron
_____ *Assg1 due, Assg2 out*
Course project proposal due
- **Week6** Project Workshop: Presentations and Peer Feedback
- **Week7** Neural Networks
_____ *Assg2 due*

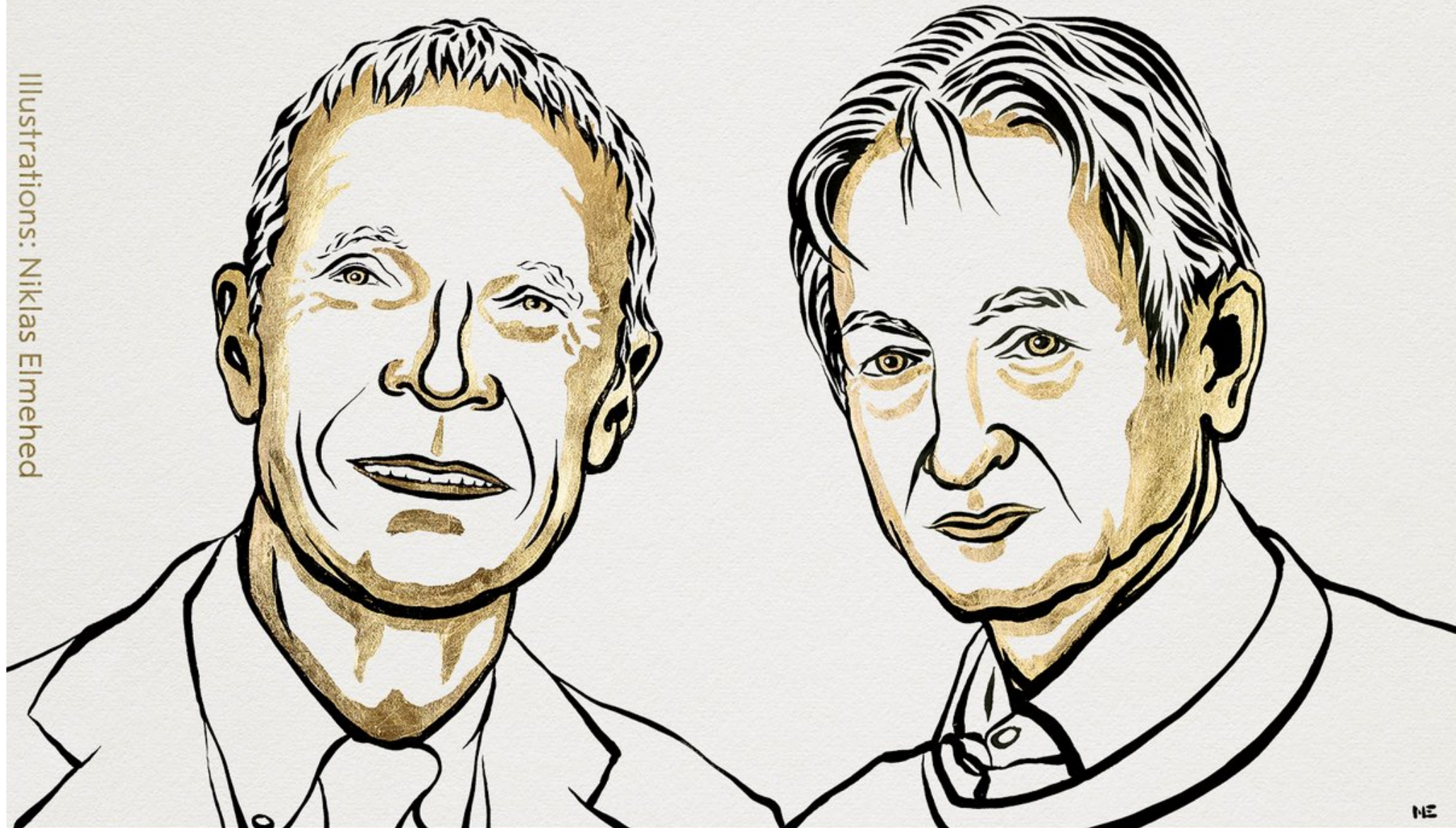
Course Outline (cont'd.)

- **Week8** Deep Learning
- **Week9** *Midterm Exam*
_____ *Assg3 out*
- **Week10** Support Vector Machines (SVMs), Multi-class SVM
- **Week11** Kernels, Support Vector Regression, Decision Tree Learning
_____ *Assg3 due*
- **Week12** Ensemble Methods: Bagging, Random Forests, Boosting
_____ *Project progress report due*
- **Week13** Clustering, Dimensionality Reduction
- **Week 14** Project Presentations
_____ *Final project report due*

Machine Learning: An Overview

THE NOBEL PRIZE IN PHYSICS 2024

Illustrations: Niklas Elmehed



John J. Hopfield

Geoffrey E. Hinton

"for foundational discoveries and inventions
that enable machine learning
with artificial neural networks"

THE ROYAL SWEDISH ACADEMY OF SCIENCES

Quotes

- *“If you were a current computer science student what area would you start studying heavily?”*
 - *Answer: Machine Learning.*
 - *“The ultimate is computers that learn”*

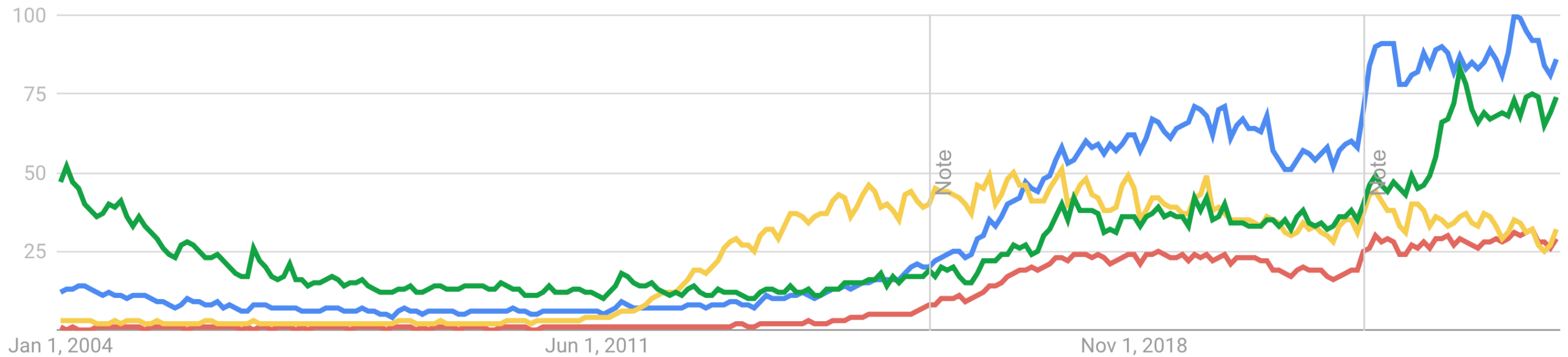
– Bill Gates, Reddit AMA
- *“Machine learning is today’s discontinuity”*

– Jerry Yang,
Co-founder, Yahoo
- *“AI is the new electricity! Electricity transformed countless industries; AI will now do the same.”*

– Andrew Ng

Google Trends

● Machine learning ● Deep learning ● Big data ● Artificial Intelligence



CORE TECHNOLOGIES

ARTIFICIAL INTELLIGENCE



DEEP LEARNING



MACHINE LEARNING



NLP PLATFORMS



PREDICTIVE APIS

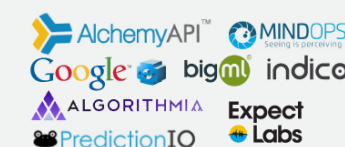


IMAGE RECOGNITION



SPEECH RECOGNITION



RETHINKING ENTERPRISE

SALES



SECURITY / AUTHENTICATION



FRAUD DETECTION



HR / RECRUITING



MARKETING



PERSONAL ASSISTANT



INTELLIGENCE TOOLS



RETHINKING INDUSTRIES

ADTECH



AGRICULTURE



EDUCATION



FINANCE



LEGAL



MANUFACTURING



MEDICAL



OIL AND GAS



MEDIA / CONTENT



CONSUMER FINANCE



PHILANTHROPIES



AUTOMOTIVE



DIAGNOSTICS



RETAIL



RETHINKING HUMANS / HCI

AUGMENTED REALITY



GESTURAL COMPUTING



ROBOTICS



EMOTIONAL RECOGNITION



SUPPORTING TECHNOLOGIES

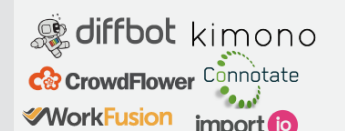
HARDWARE



DATA PREP



DATA COLLECTION



AGENTS

PROFESSIONAL	PERSONAL	OS INTERFACES

AUTONOMOUS SYSTEMS

AIR	GROUND	SEA	INDUSTRIAL

ENTERPRISE

SECURITY / FRAUD	HR / RECRUITING	SALES	MARKETING	CUSTOMER SUPPORT	INTERNAL INTEL	MARKET INTEL

PLATFORMS

RESEARCH / AGI	FULL STACK	MACHINE LEARNING	INDUSTRIAL IOT	AUDIO	VISION	DATA ENRICHMENT

INDUSTRIES

ADTECH	AGRICULTURE	FOR GOOD	RETAIL FINANCE	LEGAL	MATERIALS & MFG	HEALTHCARE

INDUSTRIES (CONT'D)

EDUCATION	TRANSPORT & LOGISTICS	INVESTMENT FINANCE	DATA SCIENCE	MACHINE LEARNING	OPEN SOURCE

ENTERPRISE INTELLIGENCE

VISUAL 	AUDIO 	SENSOR 	INTERNAL DATA 	MARKET
-------------------	------------------	-------------------	--------------------------	-------------------

ENTERPRISE FUNCTIONS

CUSTOMER SUPPORT 	SALES 	MARKETING 	SECURITY 	RECRUITING
-----------------------------	------------------	----------------------	---------------------	-----------------------

AUTONOMOUS SYSTEMS

GROUND NAVIGATION 	AERIAL 	INDUSTRIAL 	PERSONAL 	PROFESSIONAL
------------------------------	-------------------	-----------------------	---------------------	-------------------------

INDUSTRIES

AGRICULTURE 	EDUCATION 	INVESTMENT 	LEGAL 	LOGISTICS
------------------------	----------------------	-----------------------	------------------	----------------------

INDUSTRIES CONT'D

MATERIALS 	RETAIL FINANCE
----------------------	---------------------------

HEALTHCARE

PATIENT 	IMAGE 	BIOLOGICAL
--------------------	------------------	-----------------------

TECHNOLOGY STACK

AGENT ENABLERS

DATA SCIENCE

MACHINE LEARNING

NATURAL LANGUAGE

DEVELOPMENT

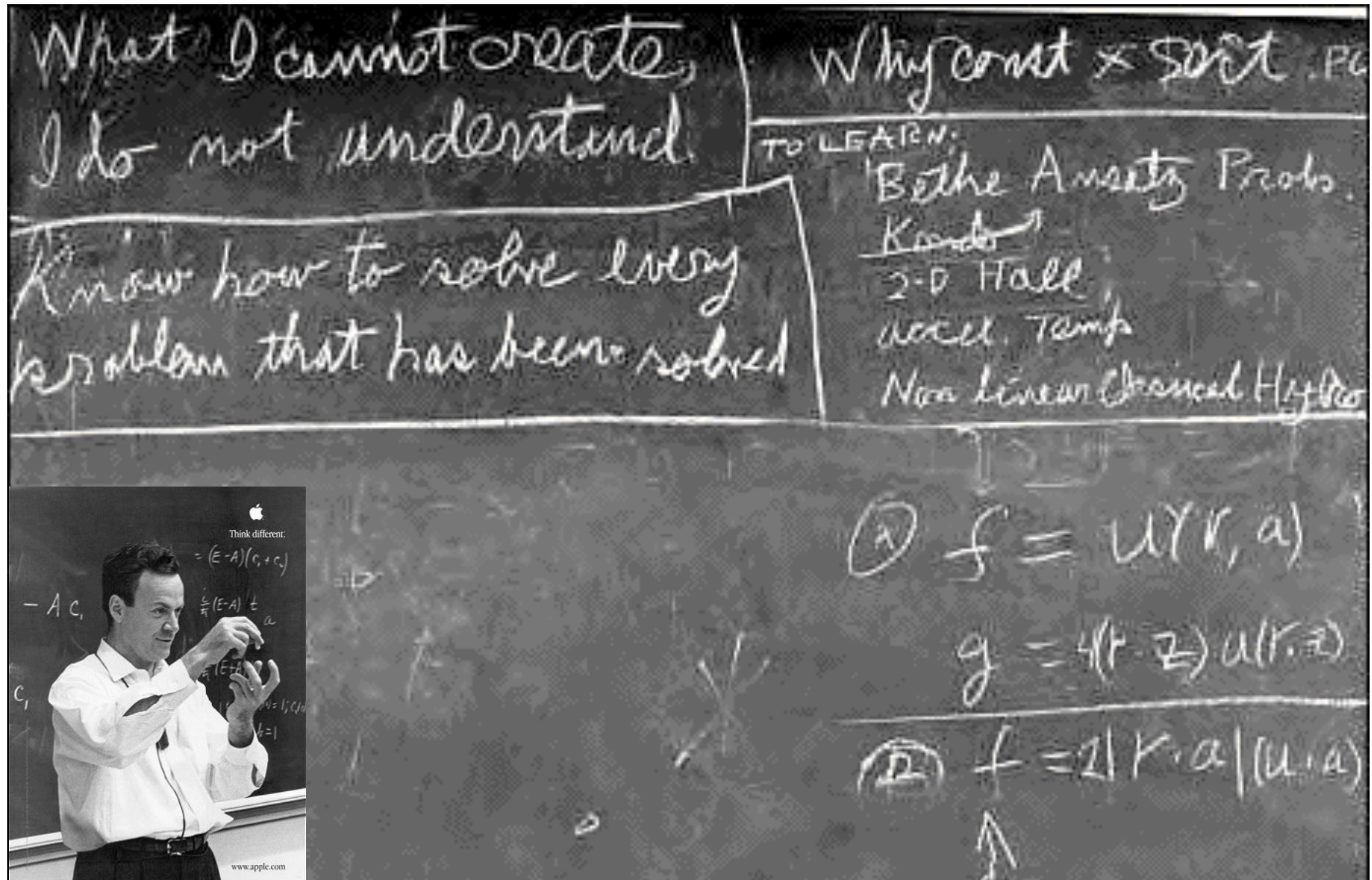
DATA CAPTURE

OPEN SOURCE LIBRARIES

HARDWARE

RESEARCH

Learning



Richard Feynman

Two definitions of learning

(1) Learning is the acquisition of knowledge about the world.

Symbolic approach *Kupfermann (1985)*

(2) Learning is an adaptive change in behavior caused by experience.

Connectionist approach *Shepherd (1988)*

Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)

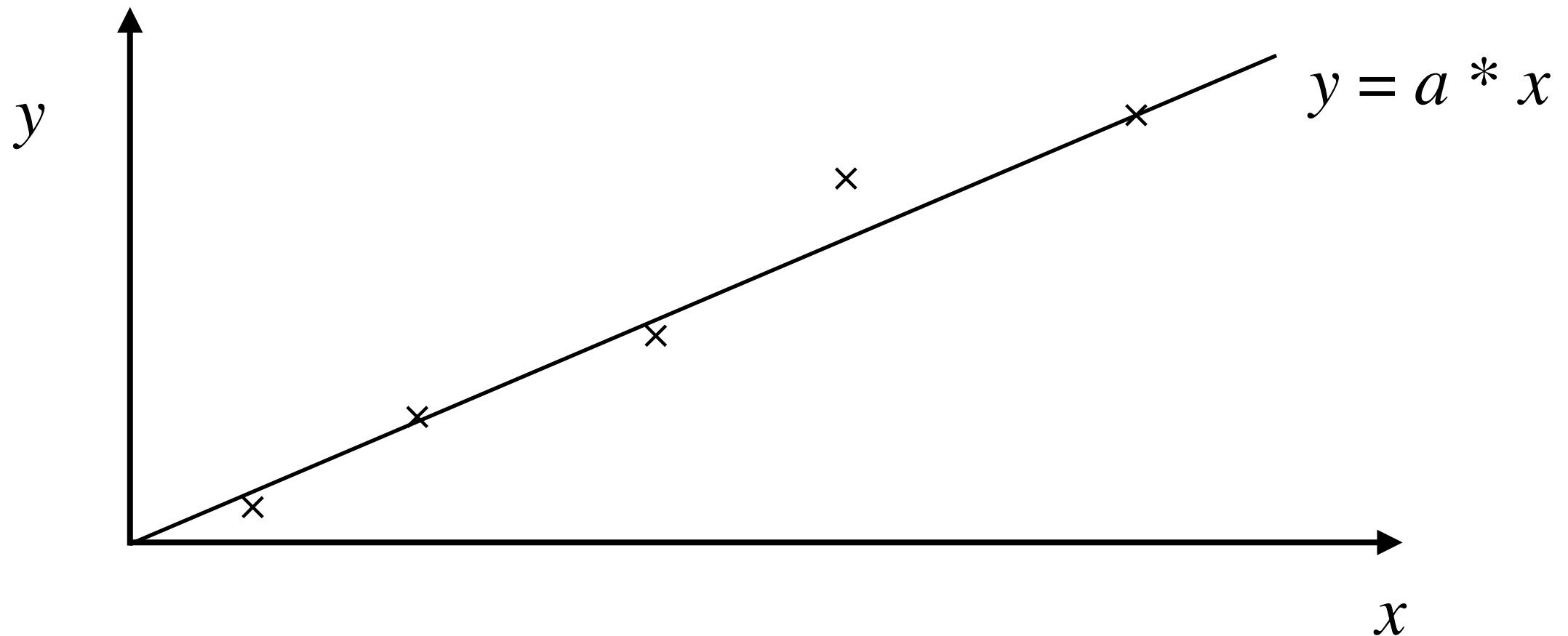
Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)
- Example 1: scientific inference



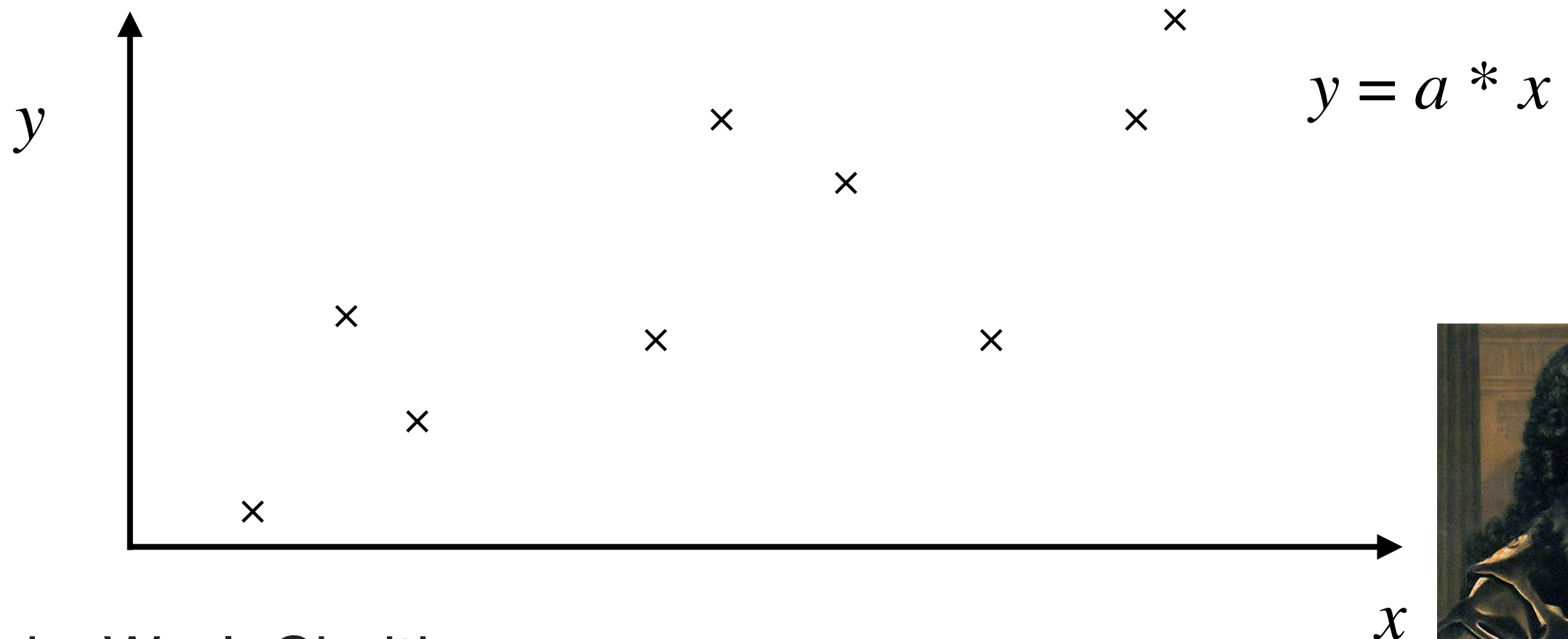
Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)
- Example 1: scientific inference



Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)
- Example 1: scientific inference

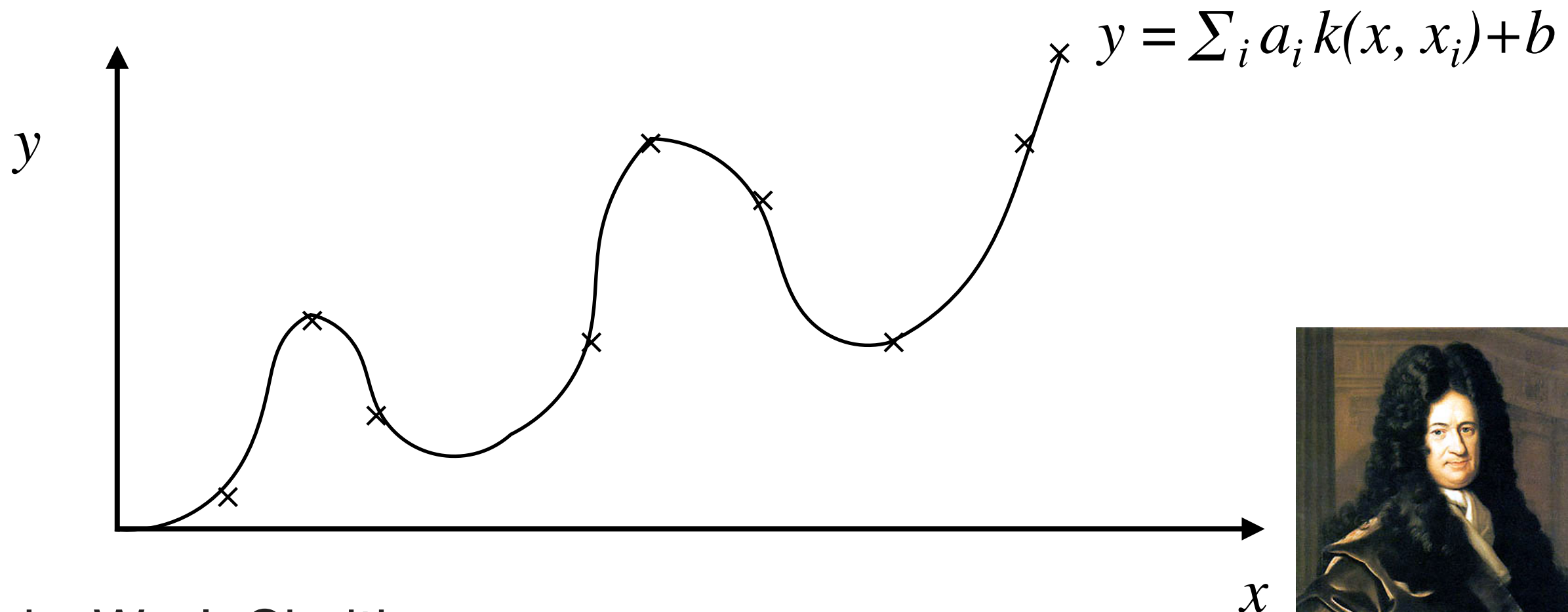


Leibniz, Weyl, Chaitin



Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)
- Example 1: scientific inference

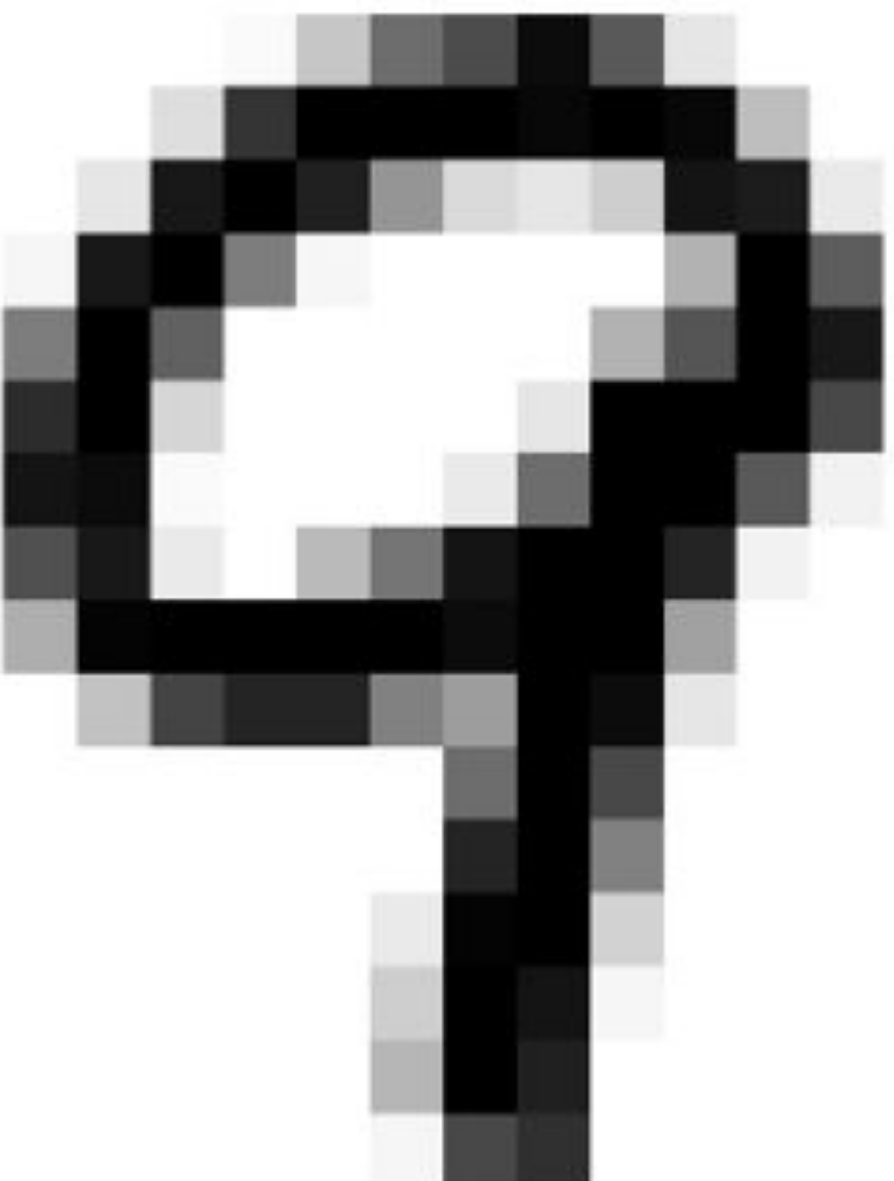


Leibniz, Weyl, Chaitin



Empirical Inference

- Example 2: perception



9



9

∞





9

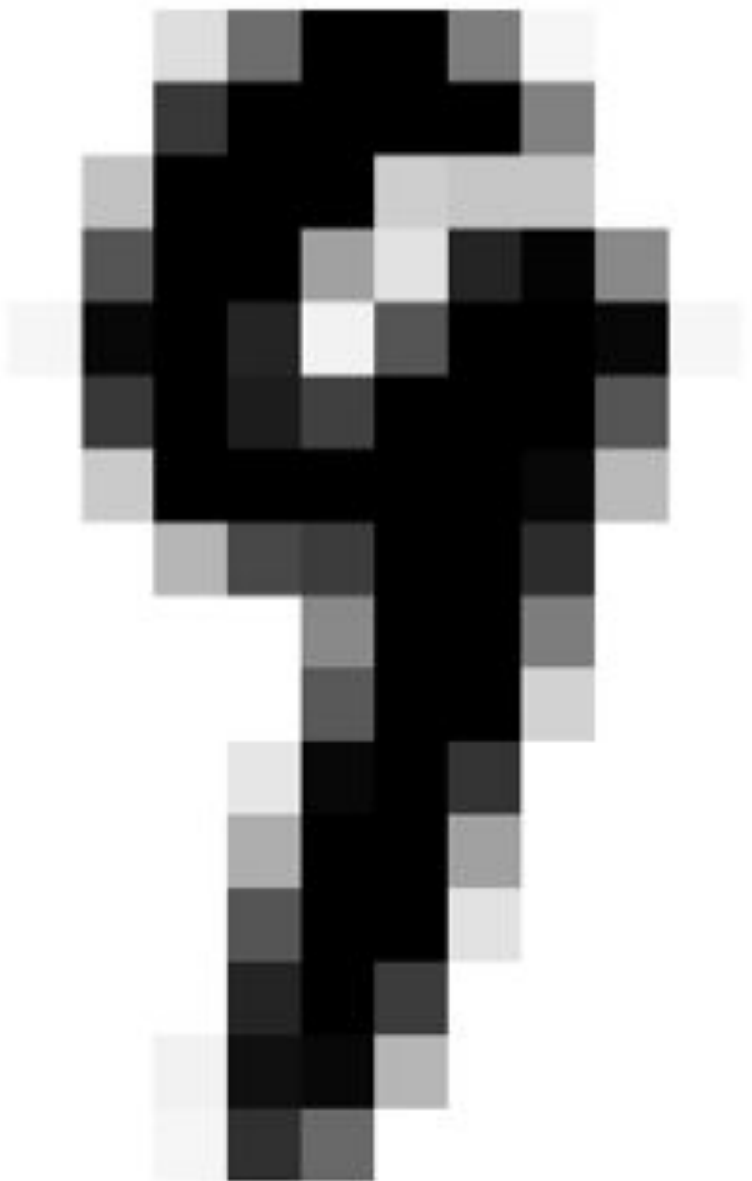


8



8

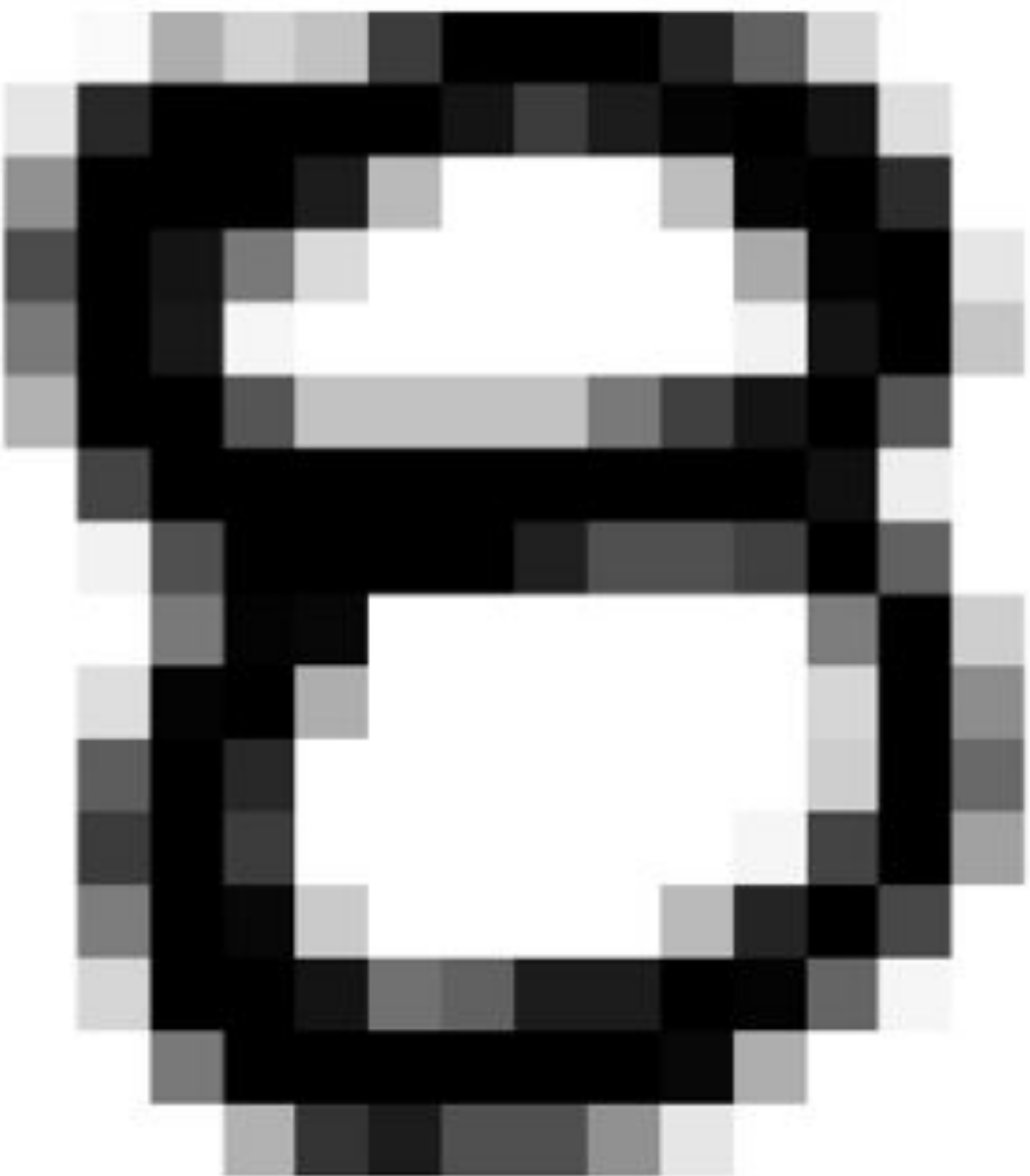
9



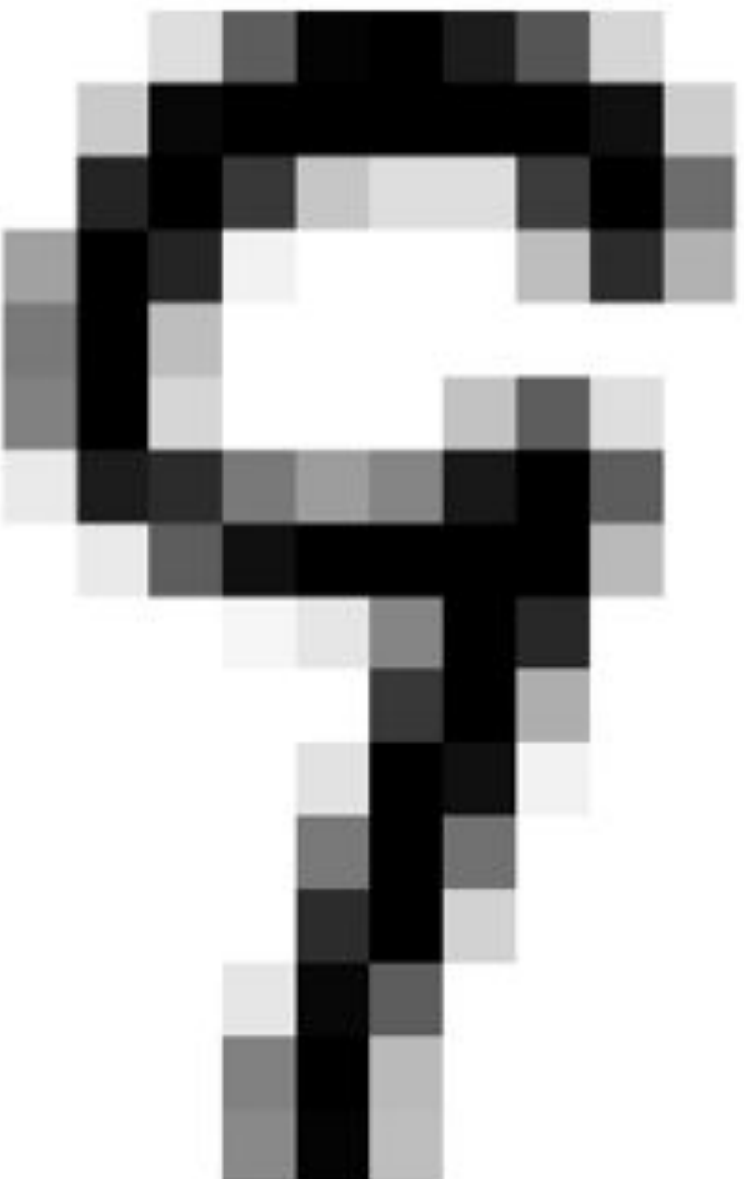
8



8



9





9



8



9



8





9



8



8



9



?



9

Empirical Inference

- Example2: perception

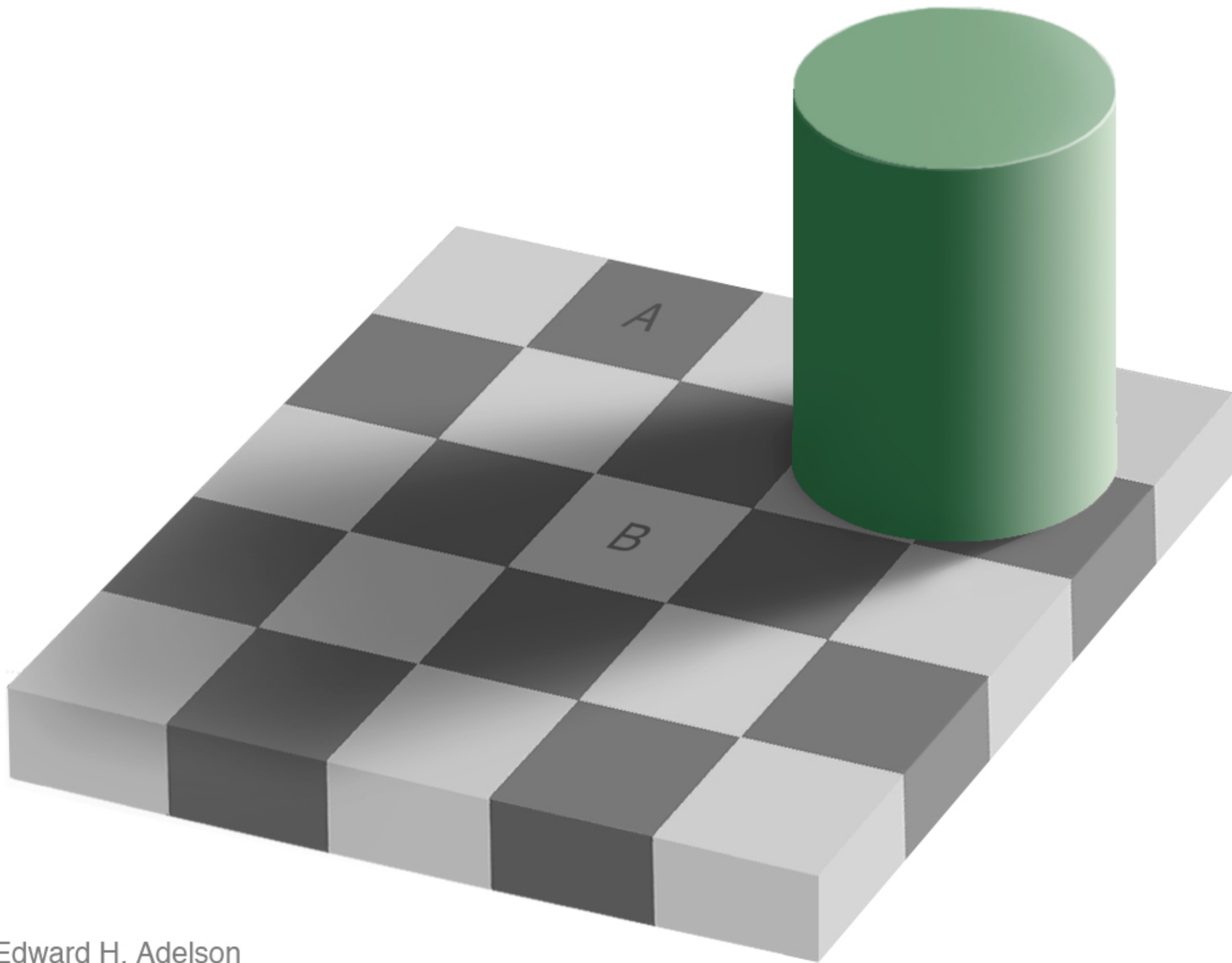
"The brain is nothing but a statistical decision organ"
H. Barlow

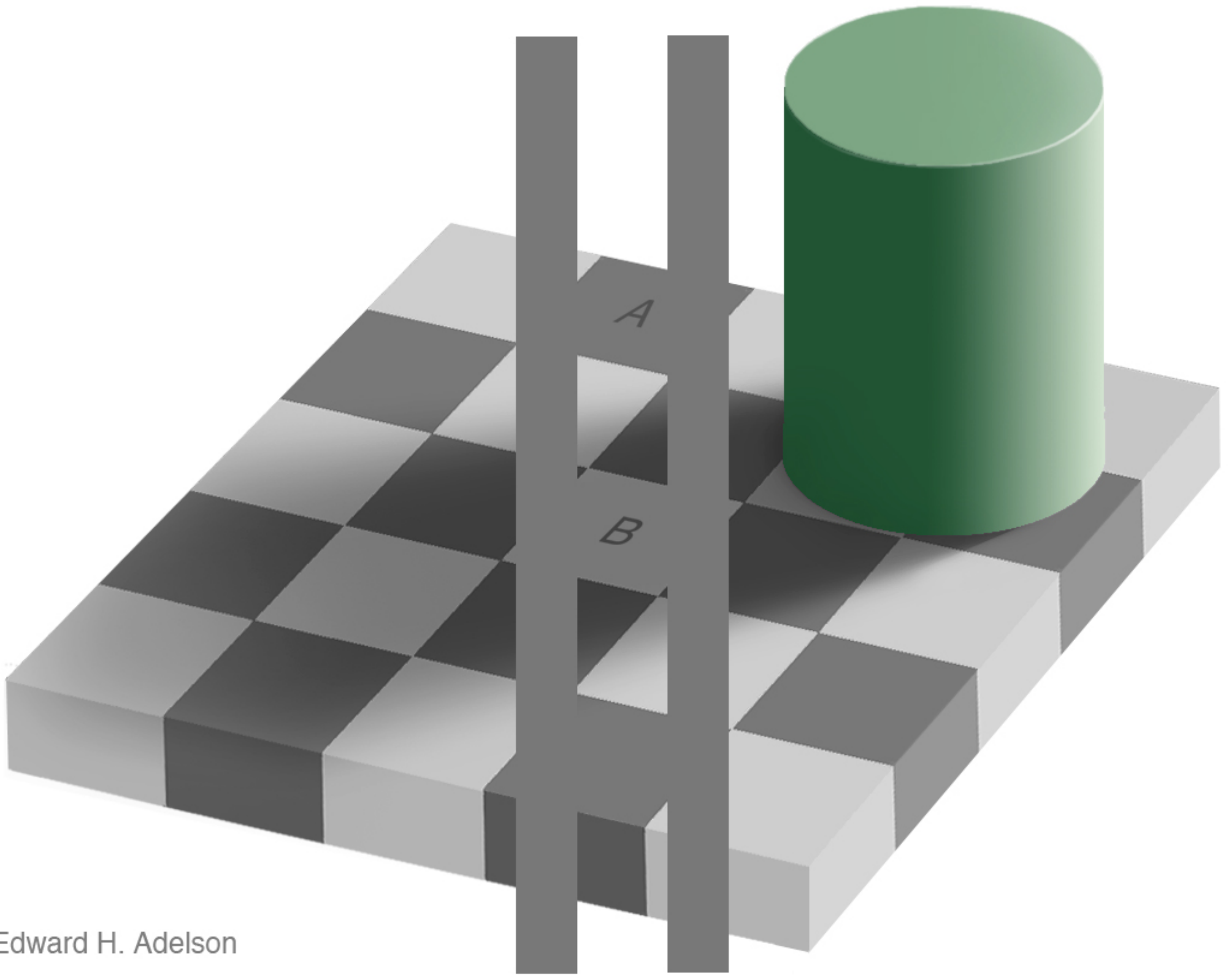
Color Perception



X

X





Edward H. Adelson

*reflected light = illumination * reflectance*

“Vision as unconscious inference”
Helmholtz

Hard Inference Problems

- High dimensionality — consider many factors simultaneously to find regularity
- Complex regularities — nonlinear; nonstationary, etc.
- Little prior knowledge — e.g. no mechanistic models for the data
- Need large data sets — processing requires computers and automatic inference methods

What is machine learning?

Example: Netflix Challenge

- Goal: Predict how a viewer will rate a movie
- 10% improvement = 1 million dollars



Example: Netflix Challenge

- Goal: Predict how a viewer will rate a movie
- 10% improvement = 1 million dollars
- Essence of Machine Learning:
 - A pattern exists
 - We cannot pin it down mathematically
 - We have data on it

AlphaGo vs Lee Sedol



NVIDIA BB8 AI Car

End to End Learning for Self-Driving Cars			
Mariusz Bojarski NVIDIA Corporation Holmdel, NJ 07735	 Davide Del Testa NVIDIA Corporation Holmdel, NJ 07735	Daniel Dworakowski NVIDIA Corporation Holmdel, NJ 07735	Bernhard Firner NVIDIA Corporation Holmdel, NJ 07735
Beat Flepp NVIDIA Corporation Holmdel, NJ 07735	Prasoon Goyal NVIDIA Corporation Holmdel, NJ 07735	Lawrence D. Jackel NVIDIA Corporation Holmdel, NJ 07735	Mathew Monfort NVIDIA Corporation Holmdel, NJ 07735
Urs Muller NVIDIA Corporation Holmdel, NJ 07735	Jiakai Zhang NVIDIA Corporation Holmdel, NJ 07735	Xin Zhang NVIDIA Corporation Holmdel, NJ 07735	Jake Zhao NVIDIA Corporation Holmdel, NJ 07735
Karol Zieba NVIDIA Corporation Holmdel, NJ 07735			

Abstract

We trained a convolutional neural network (CNN) to map raw pixels from a single front-facing camera directly to steering commands. This end-to-end approach proved surprisingly powerful. With minimum training data from humans the system learns to drive in traffic on local roads with or without lane markings and on highways. It also operates in areas with unclear visual guidance such as in parking lots and on unpaved roads.

The system automatically learns internal representations of the necessary processing steps such as detecting useful road features with only the human steering angle as the training signal. We never explicitly trained it to detect, for example, the outline of roads.

Compared to explicit decomposition of the problem, such as lane marking detection, path planning, and control, our end-to-end system optimizes all processing steps simultaneously. We argue that this will eventually lead to better performance and smaller systems. Better performance will result because the internal components self-optimize to maximize overall system performance, instead of optimizing human-selected intermediate criteria, e.g., lane detection. Such criteria understandably are selected for ease of human interpretation which doesn't automatically guarantee maximum system performance. Smaller networks are possible because the system learns to solve the problem with the minimal number of processing steps.

We used an NVIDIA DevBox and Torch 7 for training and an NVIDIA DRIVE™ PX self-driving car computer also running Torch 7 for determining where to drive. The system operates at 30 frames per second (FPS).



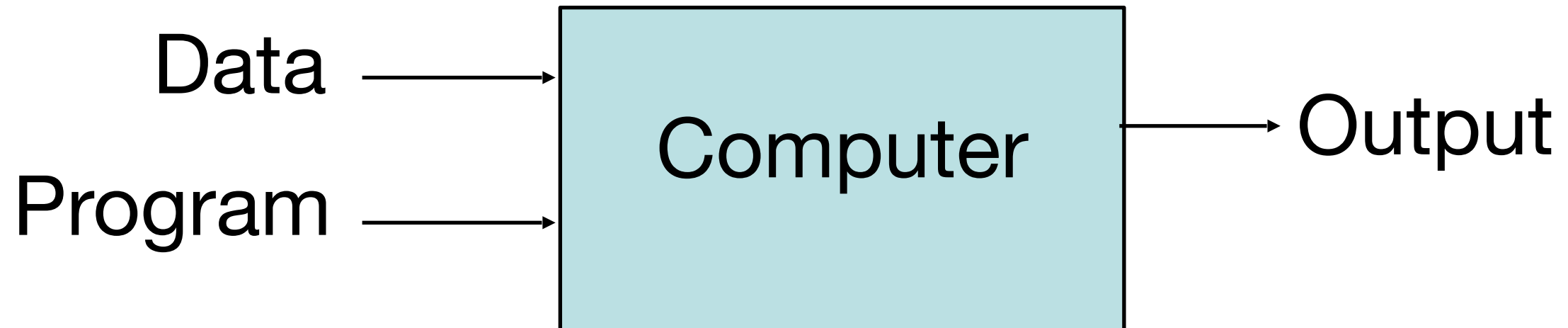
Meet NVIDIA BB8

What is Machine Learning?

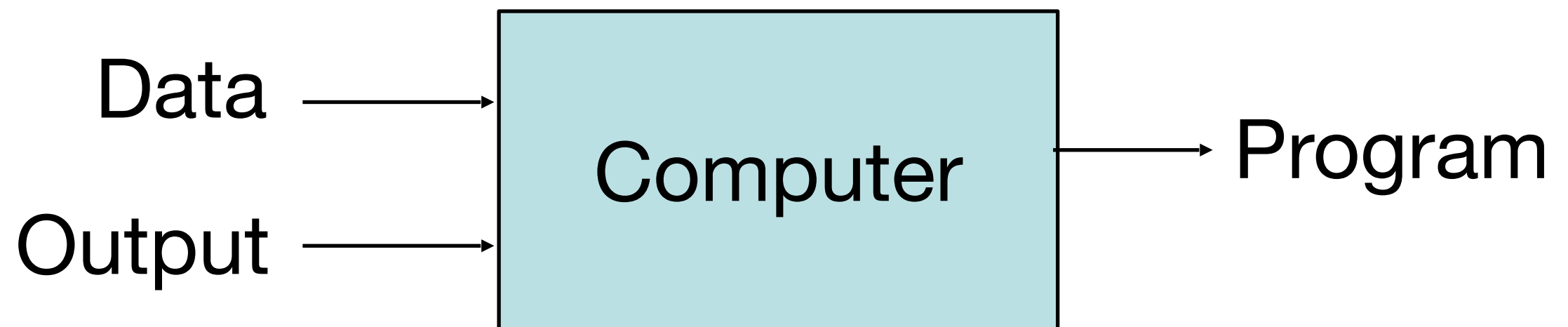
- [Arthur Samuel, 1959]
 - Field of study that gives computers the ability to learn without being explicitly programmed
- [Kevin Murphy] algorithms that
 - automatically detect patterns in data
 - use the uncovered patterns to predict future data or other outcomes of interest
- [Tom Mitchell] algorithms that
 - improve their performance (P) at some task (T) with experience (E)

Comparison

- **Traditional Programming**

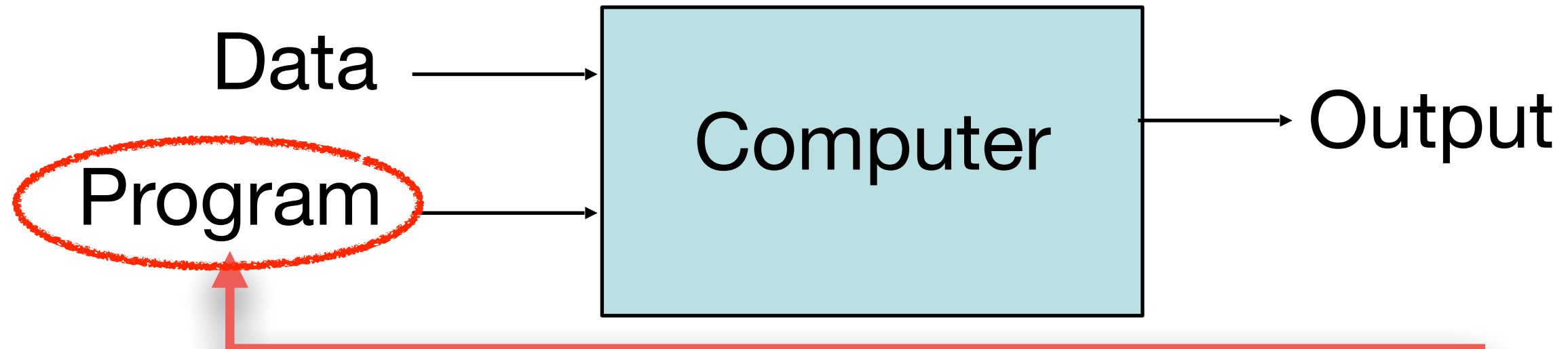


- **Machine Learning**

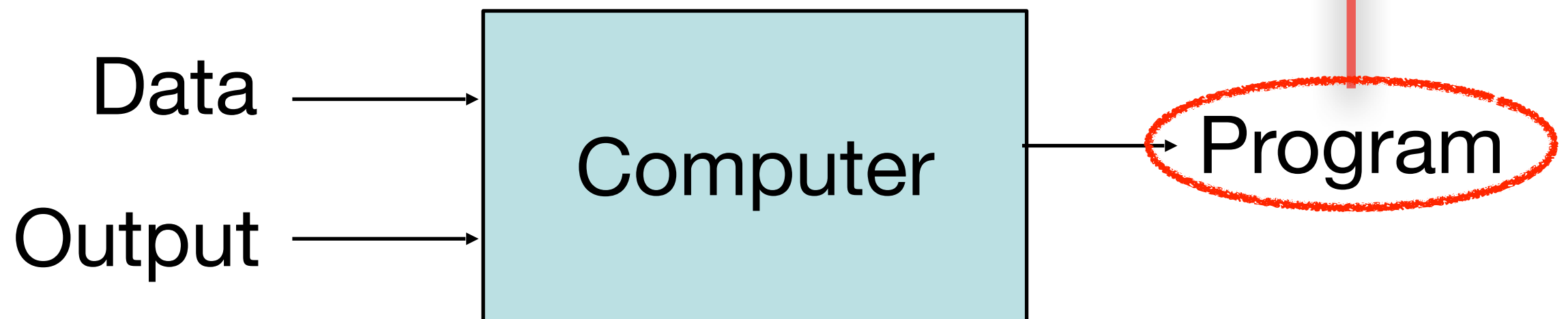


Comparison

- **Traditional Programming**



- **Machine Learning**



What is Machine Learning?

- If you are a Scientist



- If you are an Engineer / Entrepreneur
 - Get lots of data
 - Machine Learning
 - ???
 - Profit!

Why Study Machine Learning?

Engineering Better Computing Systems

- Develop systems
 - too difficult/expensive to construct manually
 - because they require specific detailed skills/knowledge
 - **knowledge engineering bottleneck**
- Develop systems
 - that adapt and customize themselves to individual users.
 - Personalized news or mail filter
 - Personalized tutoring
- Discover new knowledge from large databases
 - Medical text mining (e.g. migraines to calcium channel blockers to magnesium)
 - **data mining**

Why Study Machine Learning?

Cognitive Science

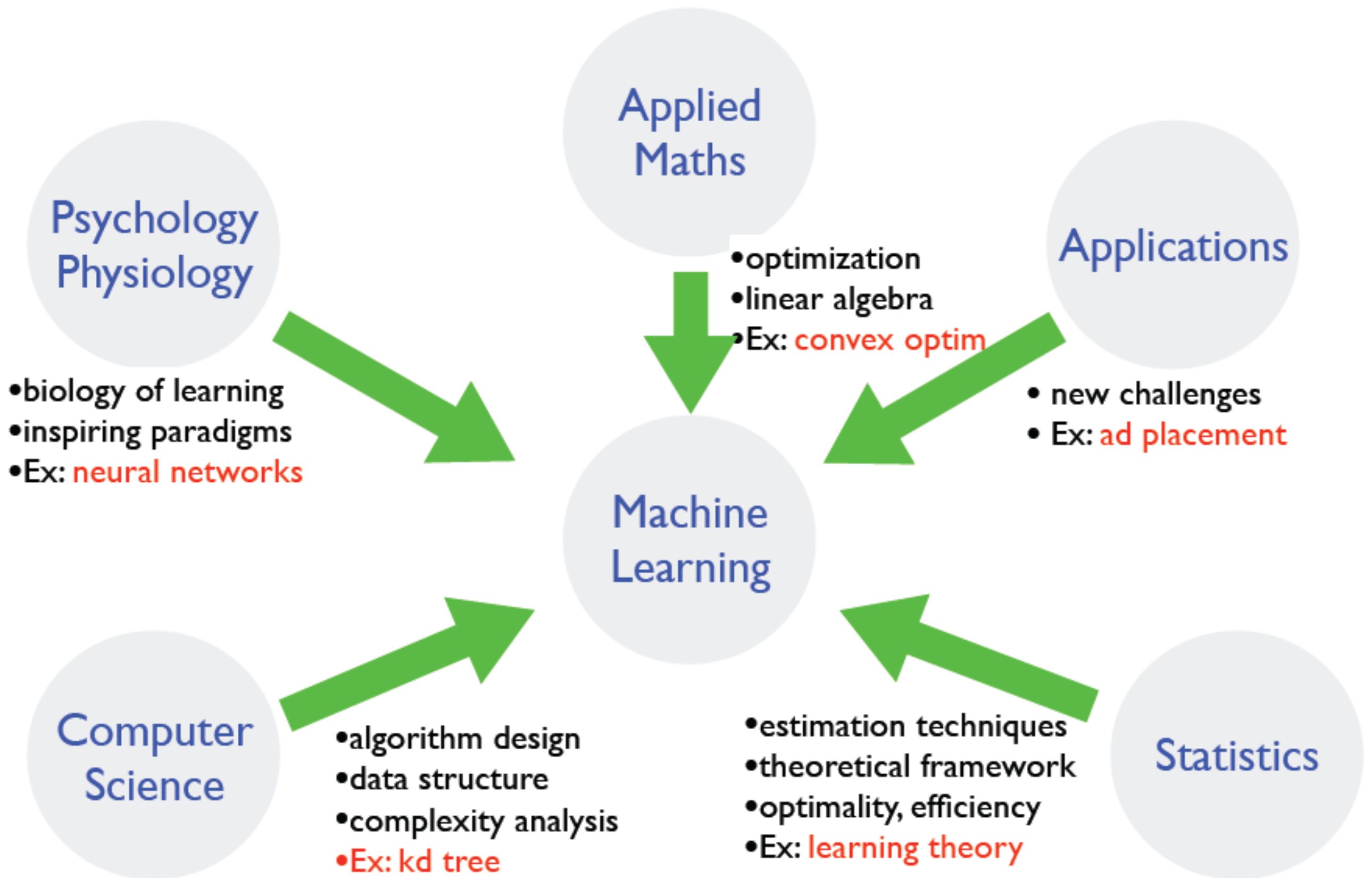
- Computational studies of learning may help us understand learning in humans
 - and other biological organisms.
- Hebbian neural learning
 - “Neurons that fire together, wire together.”

Why Study Machine Learning?

The Time is Ripe

- Algorithms
 - Many basic effective and efficient algorithms available.
- Data
 - Large amounts of on-line data available.
- Computing
 - Large amounts of computational resources available.

Where does ML fit in?



A Brief History of AI



A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.

(John McCarthy)



1956

A Proposal for the
DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

June 17 - Aug. 16

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1) Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2) How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning

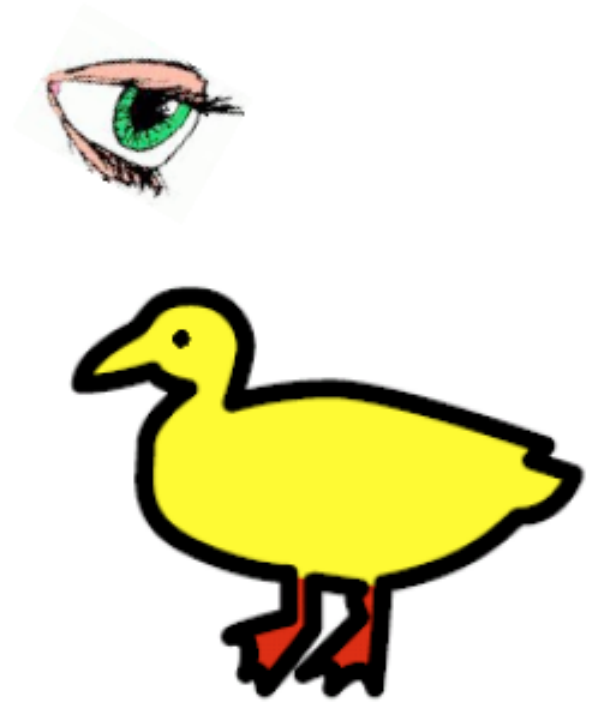


Why is AI hard?

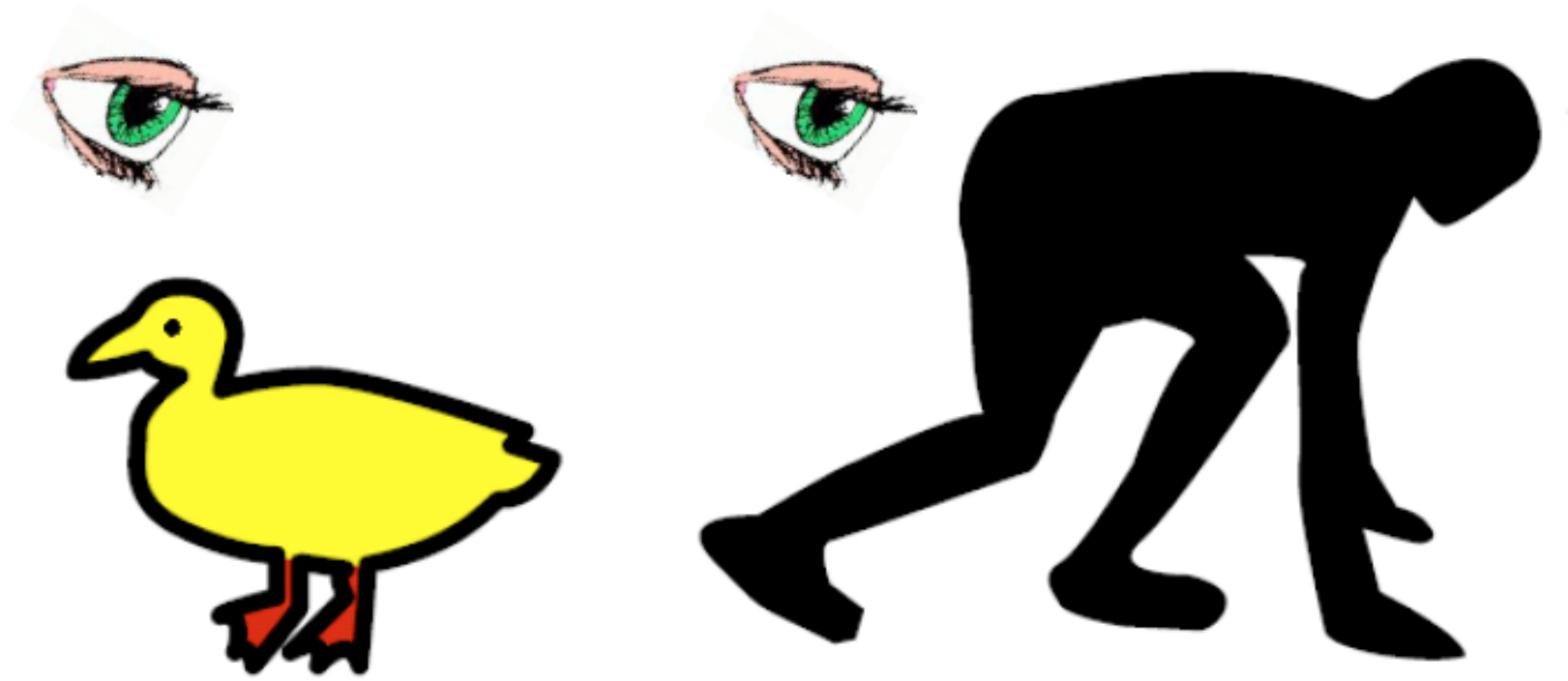
Image credit: Neşeli Günler (Arzu Film, 1978)



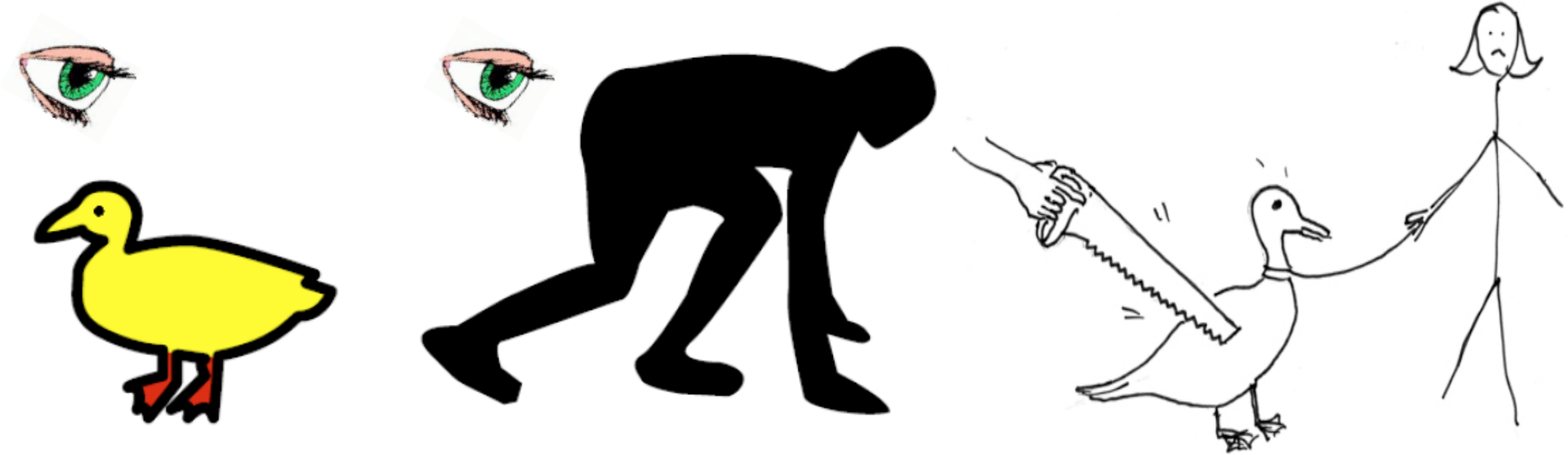
“I saw her duck”



“I saw her duck”



“I saw her duck”



Why are things working today?

- More compute power
- More data
- Better algorithms/models

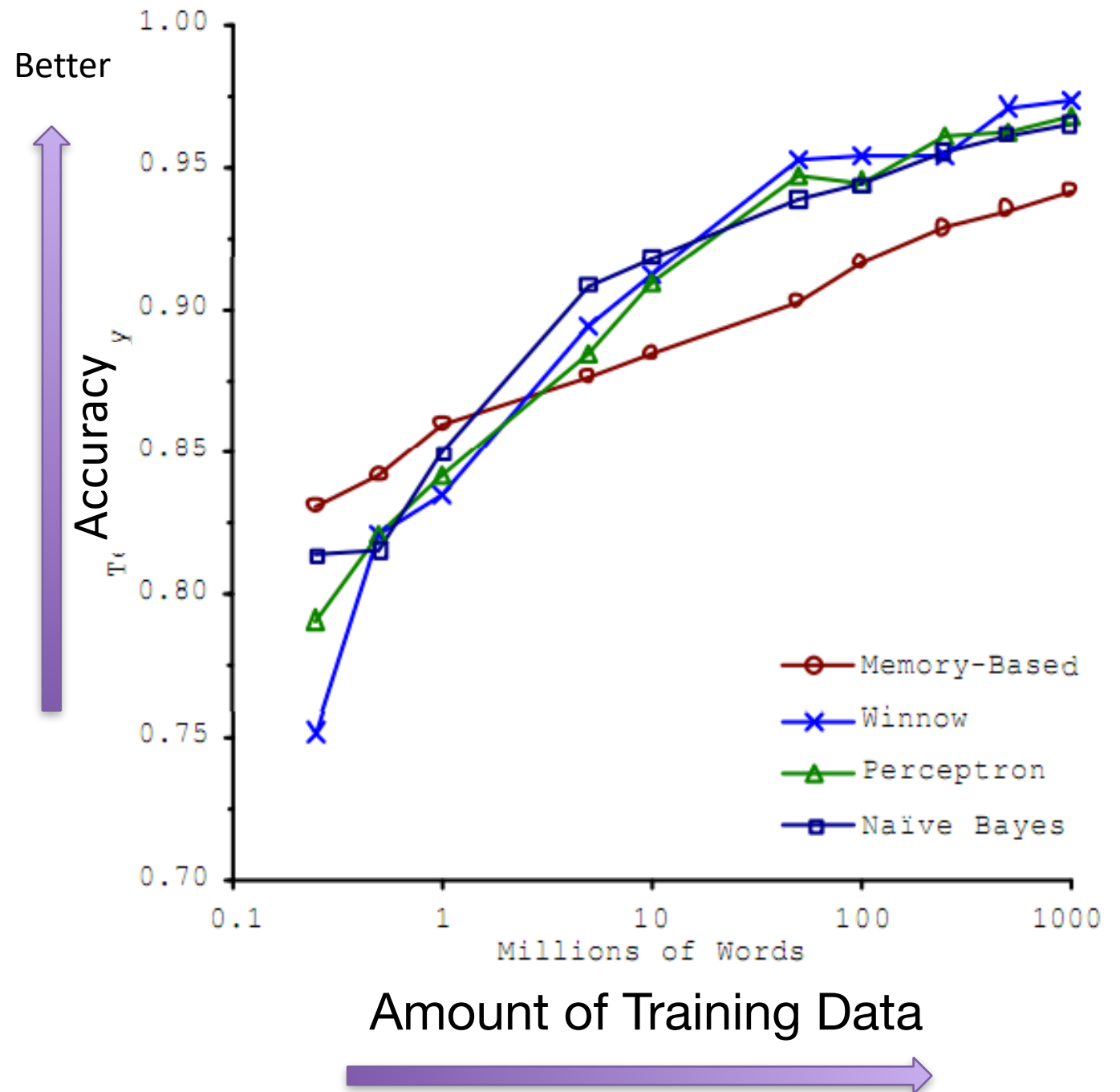


Figure Credit: Banko & Brill, 2011

Next Class:

Machine Learning by Examples,
Nearest Neighbor Classifier