Lecture 1:

- Course outline and logistics
- What is Machine Learning
Today’s Schedule

- Course outline and logistics
- An overview of Machine Learning
Course outline and logistics
Logistics

- **Instructor:**
  Aykut ERDEM  
  aykut@cs.hacettepe.edu.tr

- **Teaching Assistant:**
  Levent Karacan  
  karacan@cs.hacettepe.edu.tr
  Tugba Gurgen Erdogan  
  tugbagurgen@cs.hacettepe.edu.tr

- **Lectures:** Mon 10:00 - 11:50_D9  
  Thu 11:00 - 11:50_D9

- **Tutorials:** Fri 13:00 - 15:00_D8
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.
  – Basic linear algebra and calculus
  – Good programming skills

• BBM 409 Introduction to Machine Learning Practicum
  – Students will gain skills to apply the concepts to real world problems.
Communication

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

- We will be using Piazza for course related discussions and announcements. Please enroll the class on Piazza by following the link [http://piazza.com/class#fall2017/bbm406](http://piazza.com/class#fall2017/bbm406)
Reference Books

- A Course in Machine Learning, Hal Daumé III (online version (v.0.99) available), 2017
- Bayesian Reasoning and Machine Learning, Barber, Cambridge University Press, 2012 (online version available)
- Pattern Recognition and Machine Learning, Bishop, Springer, 2006
Grading Policy

• Grading for BBM 406 will be based on
  – a course project (done in pairs) (30%),
  – a midterm exam (30%),
  – a final exam (35%), and
  – class participation (5%)

• In BBM 409, the grading will be based on
  – a set of quizzes (20%), and
  – 3 assignments (done individually)
Assignments

• 3 assignments
  - First one worth 20%, last two worth 30% each

• **Theoretical**: Pencil-and-paper derivations
• **Programming**: Implementing Python code to solve a given real-world problem

• A quick Python tutorial in this week’s tutorial session.
KEEP CALM AND DO YOUR HOMEWORKS
Course Project

• Done individually, or in teams of two students.

• Choose your own topic and explore ways to solve the problem

  • **Proposal:** 1 page (Oct 23) (2%)
  • **Project Blogs:** Regular blog posts (4%)
  • **GitHub commits and meetings with TAs:** (4%)
  • **Progress Report:** 2-4 pages (Dec 4) (5%)
  • **Project Presentation:** Classroom presentation and video presentation (7.5%)
  • **Final Report:** 4-8 pages (Dec 29) (30%)
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.

Course Outline

- **Week 1**: Overview of Machine Learning, Nearest Neighbor Classifier
  
- **Week 2**: Linear Regression, Least Squares
  
- **Week 3**: Machine Learning Methodology
  
- **Week 4**: Statistical Estimation: MLE, MAP, Naïve Bayes Classifier

- **Assg 1 out, Assg 2 out**

- **Week 5**: Linear Classification Models: Logistic Regression, Linear Discriminant Functions, Perceptron

- **Course project proposal due**

- **Week 6**: Neural Networks

- **Assg 2 due**

- **Week 7**: Midterm Exam

- **Assg 3 out**
Course Outline (cont’d.)

• Week8  Deep Learning

• Week9  Support Vector Machines (SVMs)

• Week10 Multi-class SVM, Kernels, Support Vector Regression

• Week11 Decision Tree Learning, Ensemble Methods: Bagging, Random Forests, Boosting

• Week12 Clustering: K-Means Clustering, Spectral Clustering, Agglomerative Clustering

• Week13 Dimensionality Reduction: PCA, SVD, ICA, Autoencoders

• Week14 Course Wrap-up, Project Presentations
Machine Learning: An Overview
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 the next Internet”

  –Tony Tether, Director, DARPA

• “Machine learning is today’s discontinuity”

  –Jerry Yang, CEO, Yahoo

• “AI is the new electricity! Electricity transformed countless industries; AI will now do the same.”

  –Andrew Ng
Google Trends

- **Machine learning**
  - Field of study

- **Deep learning**
  - Field of study

- **Big data**
  - Topic

Note: The graph shows the trend of Google searches for these terms over time from Sep 1, 2007, to Jul 1, 2014.
Two definitions of learning

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

Kupfermann (1985)

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

Shepherd (1988)
Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)
Empirical Inference

• Drawing conclusions from empirical data (observations, measurements)

• Example 1: scientific inference

![Graph](image-url)
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

\[ y = \sum_i a_i k(x, x_i) + b \]
Empirical Inference

• Example 2: perception
Empirical Inference

• Example2: perception

"The brain is nothing but a statistical decision organ"

H. Barlow
Color Perception

<table>
<thead>
<tr>
<th></th>
<th>380</th>
<th>450</th>
<th>495</th>
<th>570</th>
<th>590</th>
<th>620</th>
<th>750</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td>495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td></td>
<td>570</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td>590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>620</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The image displays a spectrum of colors with wavelengths listed next to them.
reflected light = \textit{illumination} \times \textit{reflectance}
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

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**

  Data → Computer → Output

  Program → Computer

- **Machine Learning**

  Data → Computer

  Output → Program
Comparison

- Traditional Programming

  Data → Computer → Output
  Program → Computer

- Machine Learning

  Data → Computer → Program
  Output
What is Machine Learning?

• If you are a Scientist

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

Data → Machine Learning → Understanding
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?

Applied Maths
- optimization
- linear algebra
- Ex: convex optim

Psychology/Physiology
- biology of learning
- inspiring paradigms
- Ex: neural networks

Applications
- new challenges
- Ex: ad placement

Computer Science
- algorithm design
- data structure
- complexity analysis
- Ex: kd tree

Statistics
- estimation techniques
- theoretical framework
- optimality, efficiency
- Ex: learning theory
A Brief History of AI

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

(John McCarthy)
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 pre-
cisely 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 con-
sists of manipulating words according to rules of reasoning
Why is AI hard?

Image Credit: [http://karpathy.github.io/2012/10/22/state-of-computer-vision/](http://karpathy.github.io/2012/10/22/state-of-computer-vision/)
What humans see
What computers see
“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
Machine Learning
(by examples)
Pose Estimation
Collaborative Filtering

Don’t mix preferences on Netflix!

Amazon books
Collaborative Filtering

Amazon is being forced to review its website after it reportedly recommended shoppers buy items that can create explosives.

Amazon is doing some self-examination after its website suggested customers purchase potentially dangerous groupings of products.

On Wednesday, Amazon told Reuters it was "reviewing its website" after the UK’s Channel 4 News reported that the e-commerce giant’s algorithm suggests that shoppers pair certain items with products that can be used to create homemade explosives.

This chemical compound’s "frequently bought together" suggestions are the necessary ingredients to create a dangerous reaction. Amazon.com
Imitation Learning in Games

Avatar learns from your behavior

Black & White
Lionsgate Studios
Reinforcement Learning

https://www.youtube.com/watch?v=IleRKHsJBJ0
Reinforcement Learning

https://www.youtube.com/watch?v=5iZlrBqDYPM
Spam Filtering

ham

spam
Cheque Reading

segment image

Photograph Front of Check
Place the check on a dark background in a well-lit area. Keep the camera steady and align the check’s edges with the check reader.

Note: Fidelity cannot act on any written instructions

recognize handwriting
Image Layout

- Raw set of images from several cameras
- Joint layout based on image similarity
Search Ads

why these ads?
Self-Driving Cars

Image: https://medium.com/waymo/simulation-how-one-flashing-yellow-light-turns-into-thousands-of-hours-of-experience-a7a1cb475565
Speech Recognition

Given an audio waveform, robustly extract & recognize any spoken words

• Statistical models can be used to
  - Provide greater robustness to noise
  - Adapt to accent of different speakers
  - Learn from training
I need to hide a body
noun, verb, preposition, …
Face Detection

Sudhakar et al., Multi-view Face Detection Using Deep Convolutional Neural Networks, 2015
Face Detection

Yang et al., From Facial Parts Responses to Face Detection: A Deep Learning Approach, ICCV 2015
Topic Models of Text Documents
Visual Scene Understanding

- skyscrapers
- trees
- buildings
- sky
- temple
- bell
- dome
- sky

slide by Eric Sudderth
Learning - revisited

prior knowledge \rightarrow \text{Learning} \rightarrow \text{knowledge}

\text{data}
Learning - revisited

Learning

prior knowledge

Learning

knowledge

data

Crucial open problem: weak intermediate forms of knowledge that support future generalizations
Programming with Data

- Want adaptive robust and fault tolerant systems
- Rule-based implementation is (often)
  - difficult (for the programmer)
  - brittle (can miss many edge-cases)
  - becomes a nightmare to maintain explicitly
  - often doesn’t work too well (e.g. OCR)

- Usually easy to obtain examples of what we want
  IF x THEN DO y
- Collect many pairs \((x_i, y_i)\)
- Estimate function \(f\) such that \(f(x_i) = y_i\) (supervised learning)
- Detect patterns in data (unsupervised learning)
Objectives of Machine Learning

• **Algorithms:** design of efficient, accurate, and general learning algorithms to
  – deal with large-scale problems.
  – make accurate predictions (unseen examples).
  – handle a variety of different learning problems.

• **Theoretical questions:**
  – what can be learned? Under what conditions?
  – what learning guarantees can be given?
  – what is the algorithmic complexity?
Definitions and Terminology

• **Example:** an object, instance of the data used.

• **Features:** the set of attributes, often represented as a vector, associated to an example (e.g., height and weight for gender prediction).

• **Labels:** in classification, category associated to an object (e.g., positive or negative in binary classification); in regression real value.

• **Training data:** data used for training learning algorithm (often labeled data).
Definitions and Terminology (cont’d.)

- **Test data:** data used for testing learning algorithm (unlabeled data).
- **Unsupervised learning:** no labeled data.
- **Supervised learning:** uses labeled data.
- **Weakly or semi-supervised learning:** intermediate scenarios.
- **Reinforcement learning:** rewards from sequence of action.
Supervised Learning
Supervised Learning

• **Binary classification**
  Given $x$ find $y$ in $\{-1, 1\}$

• **Multicategory classification**
  Given $x$ find $y$ in $\{1, \ldots, k\}$

• **Regression**
  Given $x$ find $y$ in $\mathbb{R}$ (or $\mathbb{R}^d$)

• **Sequence annotation**
  Given sequence $x_1 \ldots x_l$ find $y_1 \ldots y_l$

• **Hierarchical Categorization (Ontology)**
  Given $x$ find a point in the hierarchy of $y$ (e.g. a tree)

• **Prediction**
  Given $x_t$ and $y_{t-1} \ldots y_1$ find $y_t$

often with loss

\[ l(y, f(x)) \]
Binary Classification
Multiclass Classification + Annotation
Regression

Linear:
- Equation: $y = 0.98x - 0.01$
- $r^2 = 0.496$
- Biological midparent beak depth (mm) vs. Midoffspring beak depth (mm)

Nonlinear:
- SPAN: 1142
- K: 0.2153
- PLATEAU: -67.44
- Graph showing a decreasing trend over time (0 to 12 minutes)
Sequence Annotation

given sequence

gene finding

speech recognition

activity segmentation

named entities

LRR Receptor-like Kinase
TIR-NBS-LRR Disease resistance
Retrotransposon associated
Other
STS
Ontology
Prediction

tomorrow’s stock price
Unsupervised Learning

Fig 3: These figures show a subset of the tree learned from the 50,000 CIFAR-100 images. The top tree only shows nodes for which there were at least 250 images. The ten shown at each node are those with the highest probability of being relevant to the node's topic.
Unsupervised Learning

- Given data \( x \), ask a good question ... about \( x \) or about model for \( x \)

- **Clustering**
  Find a set of prototypes representing the data

- **Principal Components**
  Find a subspace representing the data

- **Sequence Analysis**
  Find a latent causal sequence for observations
  - Sequence Segmentation
  - Hidden Markov Model (discrete state)
  - Kalman Filter (continuous state)

- **Hierarchical representations**

- **Independent components / dictionary learning**
  Find (small) set of factors for observation

- **Novelty detection**
  Find the odd one out
Clustering

- Documents
- Users
- Webpages
- Diseases
- Pictures
- Vehicles
...
Variance component model to account for sample structure in genome-wide association studies, Nature Genetics 2010
Sequence Analysis

Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 2007
Hierarchical Grouping

Fig 3: These figures show a subset of the tree learned from the 50,000 CIFAR-100 images. The top tree only shows nodes for which there were at least 250 images. The ten shown at each node are those with the highest probability described by a distribution over topics. In LDA, each document has a unique topic distribution. Thus multiple documents share a topic. As in latent Dirichlet allocation (LDA) [Blei et al., 2003], we consider a topic to be a distribution over words and each document to be a bag-of-words topic model, applying it to 1740 papers from NIPS 1–12.
Independent Components

find them automatically
Novelty detection

- Typical
- Atypical
Important challenges in ML

- How important is the actual learning algorithm and its tuning
- Simple versus complex algorithm
- Overfitting
- Model Selection
- Regularization