Machine Learning

What is Machine Learning?

- The real question is what is learning?
 - Using past experiences to improve future performance.
- For a machine, experiences come in the form of data.
- What does it mean to improve performance?
 - Learning is guided by an objective, associated with a particular notion of loss to be minimized (or, equivalently, gain to be maximized).
- Why machine learning?
 - We need computers to make informed decisions on new, unseen data.
 - Often it is too difficult to design a set of rules "by hand".
 - Machine learning is about automatically extracting relevant information from data and applying it to analyze new data.

Example - Visual Object Categorization





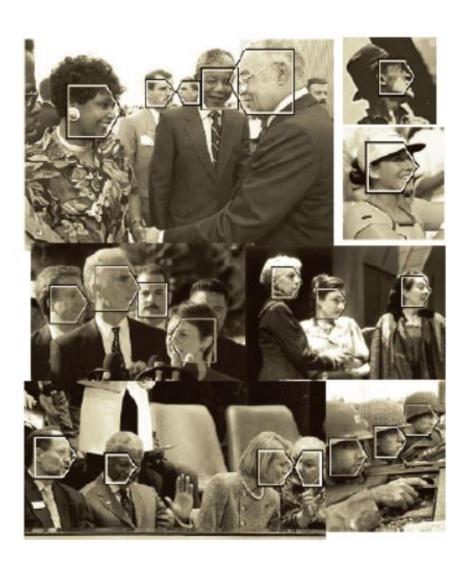
We are given categories for these images:

What are these?

- A classification problem: predict category y based on image x.
- Little chance to "hand-craft" a solution, without learning.
- Applications: robotics, HCI, web search (a real image Google...)

Example - Learning to detect objects in images

• Finding faces in images.



Example - Document Classification

- A few labeled web pages with categories: faculty, student, department, course etc.
 - Need to automatically classify previously unseen web pages.
 - What would be good features to represent these data?
- Classifying emails as spam vs not spam.

Applications of Learning

- Computer vision and robotics:
 - detection, recognition and categorization of objects
 - face recognition
 - tracking objects (rigid and articulated) in video
- Speech recognition
- Natural Language Processing
- Biology and medicine:
 - Medical outcomes analysis,
 - medical imaging and diagnosis
- Financial industry:
 - Fraud detection
 - Credit approval
 - Price and market prediction
- Information retrieval, Web search, Google ads...

What is the Learning Problem?

- Learning = Improving with experience at some task
 - Improve over task T
 - with respect to performance measure P
 - based on experience E

Supervised Learning

Supervised learning consists of the following basic steps:

Data Collection - Start with training data for which we know the correct outcome provided by a teacher or an oracle.

Representation - Choose how to represent the data.

Modeling - Choose a hypothesis class (structure of target function). This is our model of the problem.

Estimation - Find the best hypothesis.

Model Selection - We may reconsider the class of hypotheses given the outcome.

Each of these steps can make or break the learning outcome.

Some Issues in Machine Learning

- What algorithms can approximate functions well and when?
- How does number of training examples influence accuracy?
- How does complexity of hypothesis representation impact it?
- How does noisy data influence accuracy?
- What are the theoretical limits of learnability?