

Introduction to Machine Learning
Instructor: Prof. Ganesh Ramakrishnan
Lecture 1 : Introduction and Motivation

Introduction: What is Machine Learning?

- Machine learning is a sub-field of computer science that evolved from the study of **pattern recognition** and **computational learning theory** in artificial intelligence.

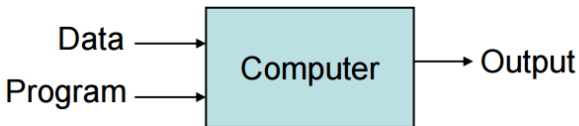
In more simpler terms:

- Using *algorithms* that iteratively learn from *data*,
- Allowing computers to find *hidden insights* without being explicitly programmed where to look

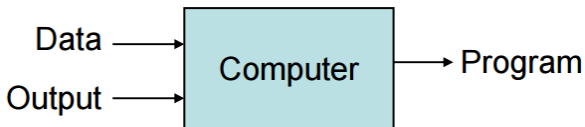
Introduction: What is Machine Learning?

- Typical algorithm has a (large) number of parameters whose values are learnt from the data
- Applications include:
 - Hand Written digit recognition
 - Face Detection
 - Spam Detection
 - Speech recognition in Google Now
 - Real-time ads on web pages and mobile devices
 -

Traditional Programming



Machine Learning

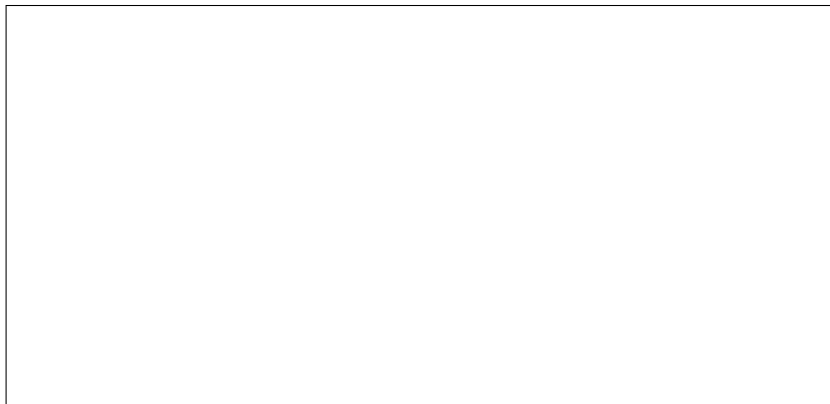


Example: Spam Detection

How to proceed...

This is an example of supervised learning problem:

- data
- training
- testing



Example: Handwritten digit recognition

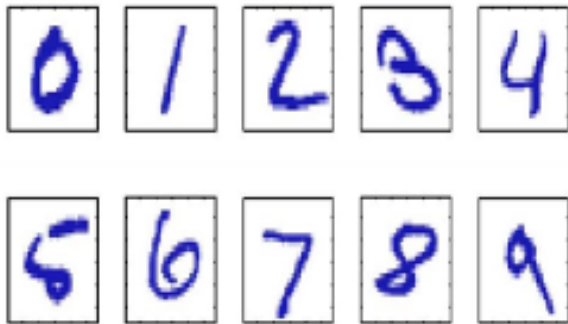


Figure: Digit recognition: Images are $28 * 28$ pixels

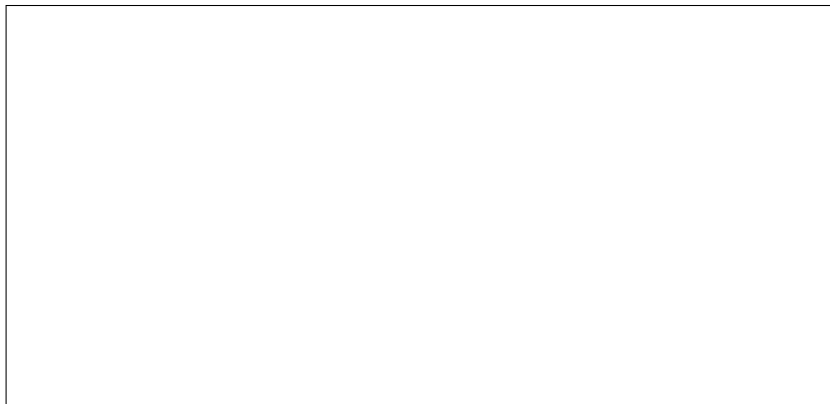
- Represent input image as a vector $x \in R^{28*28}$
- Learn a classifier $f(x)$ such that,

$$f : x \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

How to proceed...

This is an example of supervised learning problem:

- data
- training
- testing



Course Overview

- Supervised classification
 - perceptron
 - support vector machine
 - loss functions
 - kernels,
 - neural networks and deep learning
- Supervised regression
 - linear regression
 - least square linear regression model
 - Bayes Linear Regression
 - non-linear regression
 - ridge regression
 - lasso regression
 - SVM regression
- Unsupervised learning
 - clustering. K-Means
 - Expectation Maximization. Mixture of Gaussian

- **Prerequisites**

- basic Linear Algebra
- basic Probability Theory
- huge interest in learning new algorithms

- **Tutorials**

- 1 Tutorial sheet handed out every week, including a 'Tutorial 0' on the pre-requisites.
- 2 Expect students to try out each tutorial as homework
- 3 Solutions will be discussed at 1:30 PM before the following class.

- **Assignments/Homework (Individual) -**

2 assignments closely following content covered in class

- **Project - Group of 4**

Divided into 3 stages

- Stage 0 - Idea Proposals
- Stage 1 - Initial report on data-sets etc
- Stage 2 - Milestone
- Stage 3 - Final Presentation

- Quizzes
 - Quiz 1 - Week 3-4
 - Quiz 2 - Week 12
- Midsem
- Endsem

Assignments & Class Participation	20%
Quizzes	15%
Project	20%
Midsem	15%
Endsem	30%

Audit students have to attend classes, and submit assignments and project.

Notes will be periodically posted at '[cs725/calendar.html](#)' and on moodle.

Primary Book:

Elements of Statistical Learning, Trevor Hastie, Robert Tibshirani, Jerome Friedman, Springer

The following books are recommended for additional reading:

- Pattern Recognition and Machine Learning, Christopher Bishop, Springer, 2006.
 - excellent in classification and regression
- Tom Mitchell, Machine Learning. McGraw-Hill, 1997
 - good explanation of algorithms and a bible for the course
- Kevin Murphy, Statistical Machine Learning

- **Class Participation:** Every student will get points based on their participation in the following forms:
 - **Homework questions**
 - Class discussion, answering questions, asking good/foolish questions
 - Piazza/Bodhitree participation for discussing **Tutorial and Specially Marked Questions** (No private posts please!!)
 - Anything and everything which will make the course interesting

We want you to take a pledge that you will not be involved in any sort of plagiarism.

All the assignments, projects and quizzes will be checked for copy cases. In case of even a small case of copying, the name of *both the parties* will be handed over to the **DAC**¹

We also take a pledge that any sort of plagiarism will receive very strict reactions².

¹<http://www1.iitb.ac.in/newacadhome/punishments201521July.pdf>

²<http://www1.iitb.ac.in/newacadhome/procedures201521July.pdf>

Few Quotes

- A breakthrough in machine learning would be worth ten Microsofts - **Bill Gates, Chairman, Microsoft**
- Machine learning is the next Internet - **Tony Tether, Director, DARPA**
- Machine learning is the hot new thing - **John Hennessy, President, Stanford**
- Web rankings today are mostly a matter of machine learning - **Prabhakar Raghavan, Dir. Research, Yahoo**
- Machine learning is going to result in a real revolution - **Greg Papadopoulos, CTO, Sun**
- Machine learning is today's discontinuity - **Jerry Yang, CEO, Yahoo**