Introduction to Machine Learning - CS403/725
Instructor: Prof. Ganesh Ramakrishnan
Lecture 1 : Introduction and Motivation

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

Handwritten annotations:

① Hardness of learning
② Hardness of predictions
③ Associating confidence

Consistency in observation & expectation

Parameters (language of PR)

→ Eg: Stochastic # optimization in Neural Network

lead to new actions

Predicting strength of CS725

① BTech & MTech strength (increase)

② (Increase in the) number of courses

③ Grading statistics (performance of students)
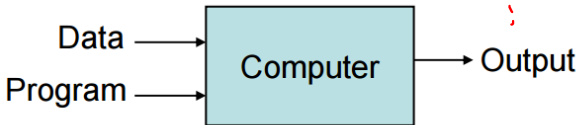
④ Periodicity

⑤ Number of students already taken course in prev sem
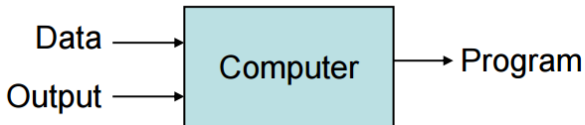
⑥ Performance of instructor

Introduction: What is Machine Learning?

- Typical algorithm has a (large) number of parameters whose values are learnt from the data

- Application includes:
  - Hand Written digit recognition
  - Face Detection
  - Spam Detection
  - Speech recognition in Google Now
  - Real-time ads on web pages and mobile devices
  - ....

Traditional Programming

Data ⟶

Program ⟶

Computer ⟶ Output

Human feedback

Machine Learning

Data ⟶

Output ⟶

Computer ⟶ Program

Pattern 1: If email has you & does not have a name

then spam

Rules (Decision lists & decision trees):

else it is a ham

Pattern 2: If 0.6 [# of "you"] $- 0.4$[# of names] exceeds 0.1 then spam
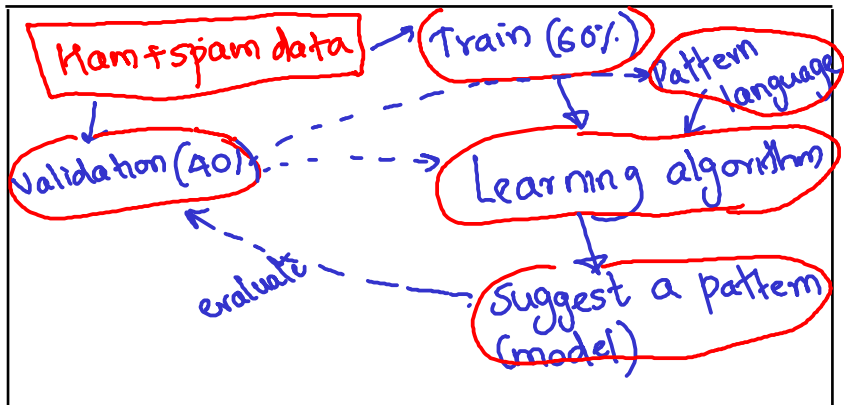
Linear classifiers (Sum, LM)

Other examples: Polynomial combinations

This is an example of supervised learning problem:

- data
- training
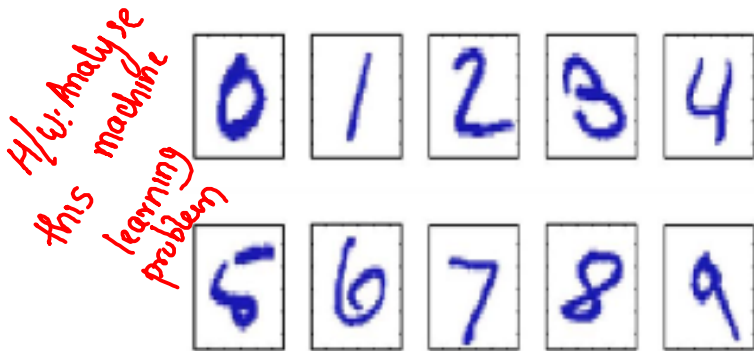- testing

Example: Handwritten digit recognition

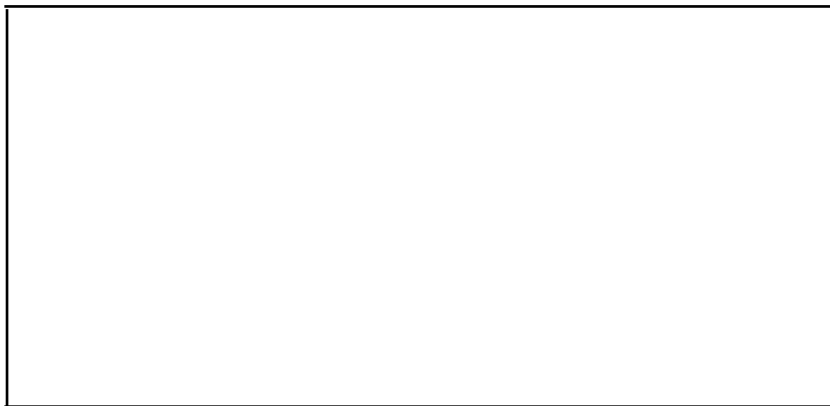Figure: Digit recognition: Images are $28 * 28$ pixels

- Represent input image as a vector $x \in R^{28*28}$
- Learn a classifier $f(x)$ such that,
$$f : x \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

H/w: Analyse this machine learning problem

This is an example of supervised learning problem:

- data
- training
- testing

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

    There will be tutorials on the prerequisite of the course and a quiz at the end of week 1 to test the knowledge

- **Assignments/Homework (Individual)** -
  2 assignments closely following content covered in class room

- **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	20%
Quizzes	15%
Project	20%
Midsem	15%
Endsem	30%

Course Materials

Notes will be periodically posted on ~~piazza~~ *tutorspace* and moodle.

Primary Book:

Elements of Statistical Learning, Trevor Hastie, Robert Tibshirani, Jerome Friedman, Springer
The following books are recommended for 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

Classroom Policy

- **Houses:** Class will be divided into 5 groups/houses. Every student will get points based on their participation in the following forms:
  - Class discussion, answering questions, asking good/foolish questions
  - ~~Piazza~~ participation (No private posts please!!) *Tutor space* →
  - Homework questions
  - Anything and everything which will make the course interesting

- **Rewards:** Winning house will get a reward and top 10 scorers will be personally treated by the instructor. And there is much more :)

- **TA Speak:** Every two week there will be presentation from a TA on some state-of-the-art technologies/applications of machine learning, which will help broaden your knowledge about the field.

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 given to the **DAC**

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

- 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**