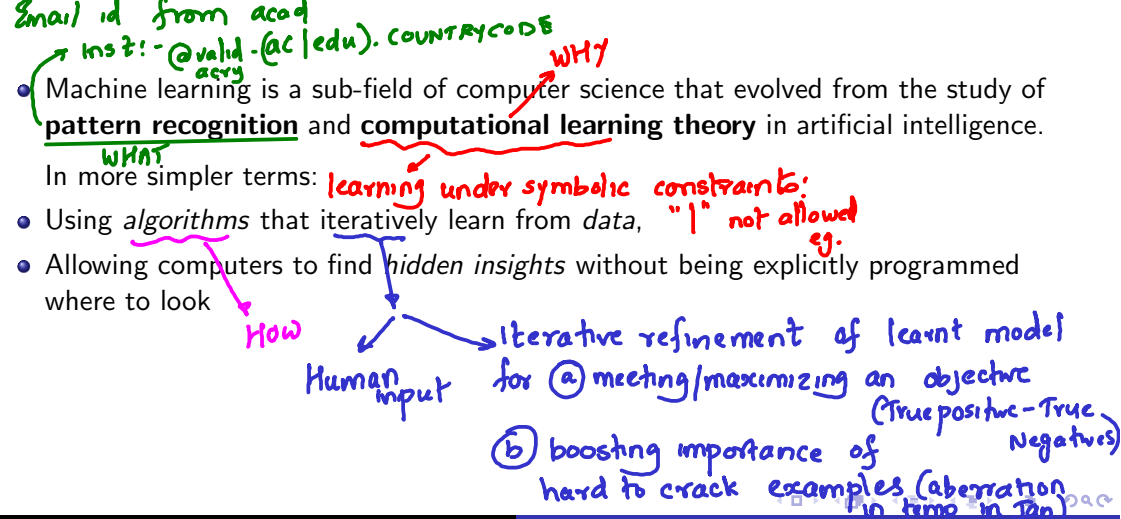
Introduction to Machine Learning
Introduction and Motivation

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.

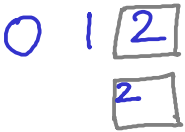
  In more simpler terms:

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

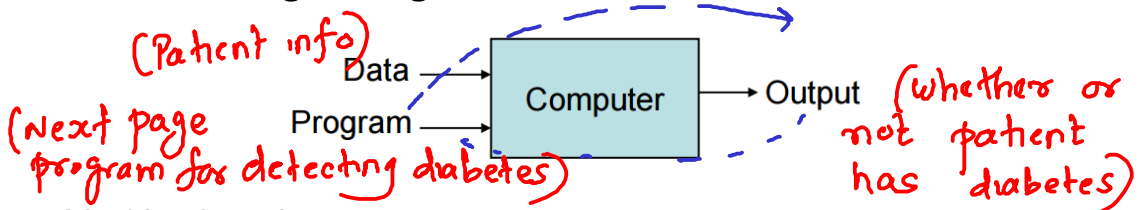
Introduction: What is Machine Learning?

desirable: {param set} should be invariant to translation, scaling, some rotation, blur

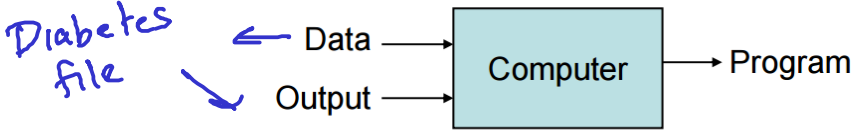
- Typical algorithm has a (large) number of parameters whose values are learnt from the data
- Applications include:
  - Hand Written digit recognition
  - Face Detection
  - Spam Detection
  - Speech recognition in Google Now
  - Real-time ads on web pages and mobile devices
  - ....



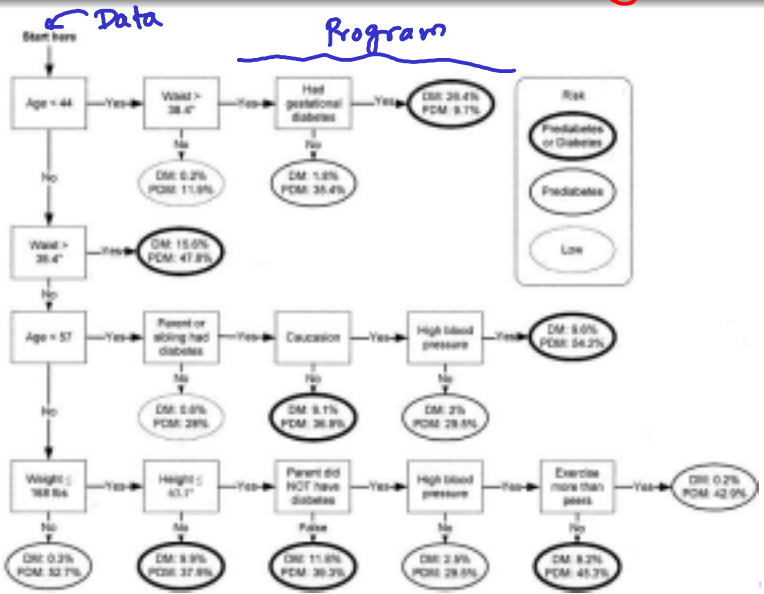
Traditional Programming



Machine Learning



Example: Classification for Diabetes (Conventional Programming)



```
@relation pima_diabetes
@attribute 'preg' real
@attribute 'plas' real
@attribute 'pres' real
@attribute 'skin' real
@attribute 'insu' real
@attribute 'mass' real
@attribute 'pedi' real
@attribute 'age' real
@attribute 'class' { tested_negative, tested_positive}
@data
6,148,72,35,0,33.6,0.627,50,tested_positive
1,85,66,29,0,26.6,0.351,31,tested_negative
8,183,64,0,0,23.3,0.672,32,tested_positive
1,89,66,23,94,28.1,0.167,21,tested_negative
0,137,40,35,168,43.1,2.288,33,tested_positive
```

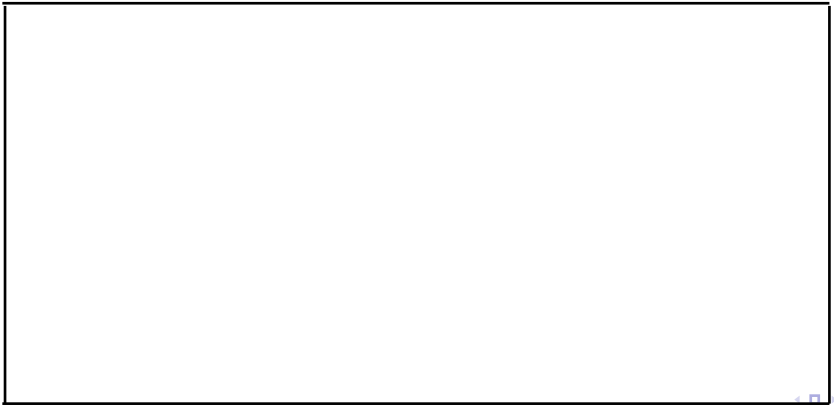
Example: What to do with a Diabetic Patient!

(Multi-class classification)

How to proceed...

This is an example of supervised learning problem:

- data
- training
- testing



Example: Handwritten digit recognition

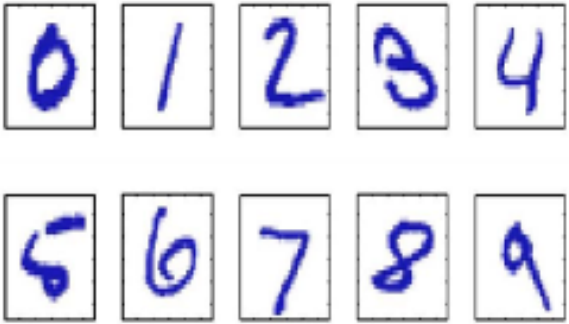


Figure: Digit recognition: Images are $28 * 28$ pixels

piecewise combination of a quadratic & 2 lines

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

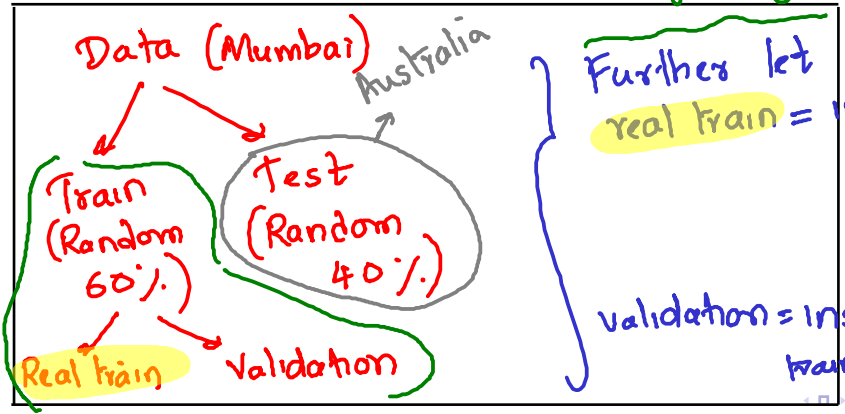
$$f : x \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

How to proceed...

This is an example of supervised learning problem:

- data → rainfall prediction between [1947...2016]
- training
- testing

Design a good Ranjit trophy to beat Australia.



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

Logistics

- **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.

Course Materials

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
- Deep Learning Book Joshua Bengio

Good for intuitive feel

An excellent read
Very good at prob models

NO: Tree learning, unsupervised
= Deep

Classroom Policy

- **Flipped Classroom**
- **Class Participation:** Every student will get points based on their participation in the following forms:
  - **Quiz based on previous class at the beginning**
  - Class discussion, answering questions, asking good/foolish questions
  - Participation (through LMS) for discussing **Tutorial and Specially Marked Questions** (No private posts please!!)
  - Anything and everything which will make the course interesting

Need to have watched video lectures for that class. Class discussions around running examples.

Honor Code

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 reactions[2].

[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**

Introduction to Machine Learning
Supervised vs. Unsupervised Learning and Method of Least Squares

Supervised vs Unsupervised

Task: Suppose you had a basket and it is fulled with some fresh fruits your task is to arrange the same type fruits at one place.
Suppose the fruits are apple, banana, cherry, grape

Case: 1

- You already know: Shape (parametrize shape?), Color
- **Train data**: Pre-classified data
- Goal: Learn from the pre-classified data and predict on new unclassified fruits.
- This type of learning is called as **supervised learning**.

Case 2:

- In this case, you know nothing about the fruits, you are seeing them for the first time!
- How will you arrange fruits of the same type together?
- One approach is to consider various characteristics of a fruit and divide them on the basis of that.
- Suppose you divide the fruits on the basis of *color* first.
  - ... Apple, cherry,
  - ... Custard apple, pear, grape
- Now you take another physical characteristic, size. The grouping will then be:
  - ... Apple, Pear, Custard apple
  - ...
  - .. grape, cherry
  - ...
- ..

Case 2:

- In this case, you know nothing about the fruits, you are seeing them for the first time!
- How will you arrange fruits of the same type together?
- One approach is to consider various characteristics of a fruit and divide them on the basis of that.
- Suppose you divide the fruits on the basis of *color* first.
  - **Red Color Group**: Apples and cheery
  - **Green Color Group**: Bananas and grapes
- Now you take another physical characteristic, size. The grouping will then be:
  - **Red color and big size**: Apple
  - **Red color and small size**: Cheery
  - **Green color and big Size**: Banana
  - **Green color and small Size**: Grapes

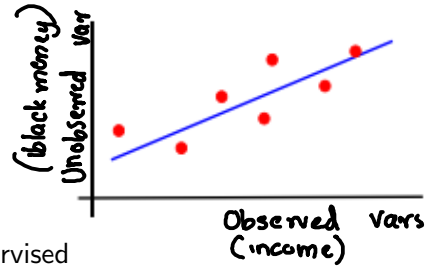
} Iterative refining the space

- This type of learning is **unsupervised learning**

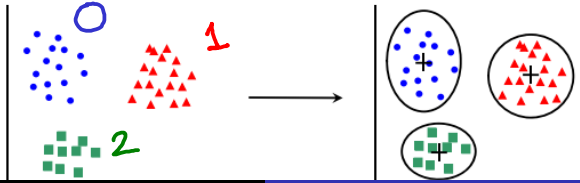
- In supervised learning, the desired outputs are provided which are used to train the machine whereas in unsupervised learning no desired outputs are provided, instead the data is analysed and studied through clustering, mining associations, reduce dimensionality, *etc.* into different classes

Three Canonical Learning Problems

1. Regression - Supervised
   - Estimate parameters, e.g. least square fit

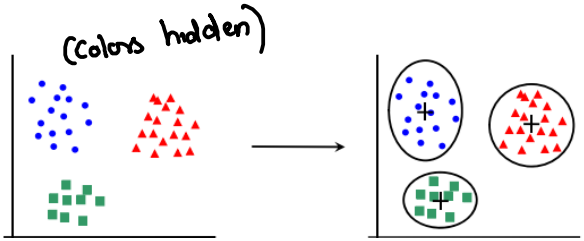


2. Classification - Supervised
   - estimate class, eq handwritten digit classification

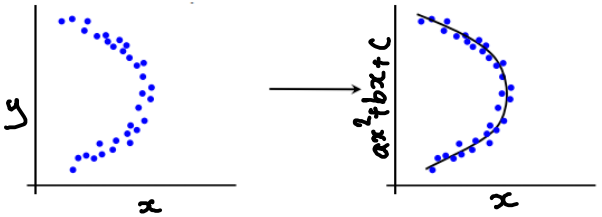


3 Unsupervised Learning - model the data

- clustering



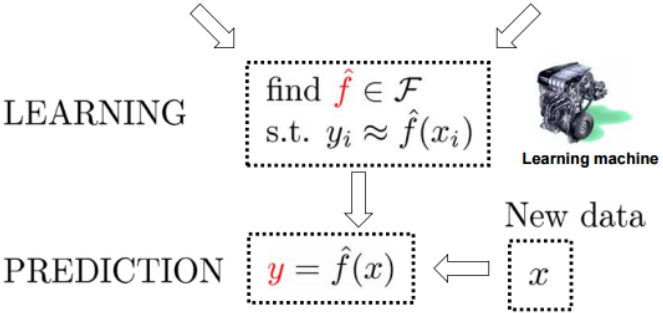
- dimentionality reduction



Supervised Learning

Functions F Training Data

$f : X \to Y \quad \{ (x^i, y^i) \in X * Y \}$



Next ....

We will start with linear regression and least square method to calculate parameters for linear regression problems.

Recap

- **Machine Learning in general**
  - Supervised Learning
  - Unsupervised Learning
  - Applications and examples
- **Canonical Learning Problems**
  - Regression Supervised
  - Classification Supervised
  - Unsupervised modeling of data

Agenda

- What is data?
  - Noise in data
- How to predict?
  - Fitting a curve
  - Error measurement
  - Minimizing Error
- Method of Least Squares

What is data?

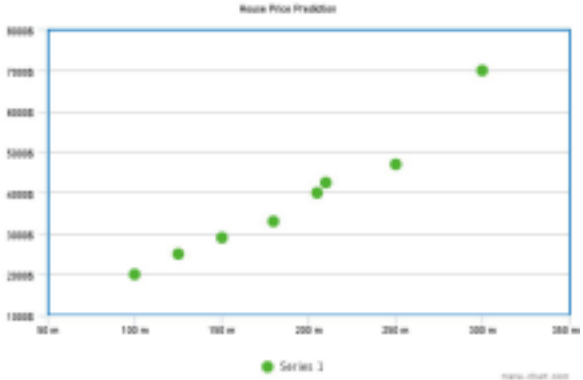
- For us, data is the information about the problem, you are solving using ML, in quantized form
- This data can be from any source, some examples are
  - Prices of stock and stock indexes such as BSE or Nifty
  - Prices of house, area and size of the house
  - Temperature of a place, latitude, longitude and time of year
- The objective of ML is to predict or classify something using the given data
- Hence, one or more than one parameters of the data must also represent the output of our program

Noise in Data

- Data in real life problems are generally collected through surveys
- And surveys may have random human errors
- Hence most methods we will be using deals with expectations as they minimize the effect of error in our predictions
- It is better to find outliers and clean data in the first step. This is known as data cleansing

Example dataset for this lecture

- For this lecture we will consider variation of cost of the house with the area of the house
- In this example we want to find a pattern or curve which this dataset follows, hence predict the price for any value of area



How to predict?

- Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. - Wikipedia
- Thus we need a critera to compare two curves on a dataset
- We describe an error function F(f, D) which takes a curve f and dataset D as input and returns a real number
- Error function must be such that it can capture how worse is our

Example

- Consider the example below where we have two curves on our dataset defined by blue(f_b) and red(f_r) line respectively. We want to find which is the better fit.

Figure: House purchase data curve fit

Question

What are some options for F(f,D)?
Hint: Measurement of difference from original value.

Examples of F

- $\sum_D f(x_i) - y_i$
- $\sum_D |f(x_i) - y_i|$
- $\sum_D (f(x_i) - y_i)^2$
- $\sum_D (f(x_i) - y_i)^3$
- and many more

Question

What F do you think can give us best fit curve and why?
Hint: Intuition of distances.

Squared Error

$$\sum_{D}(f(x_i) - y_i)^2$$

- To find the best fit curve we try to minimize the above function
- It is continuous and differentiable
- It can ve visualized as square of Euclidean distance between predicted points and actual points
- How we can perform mathematical treatment over this function will be covered in further lectures.
- This mathematical treatment is known as method of least squares. Can you find the reason why it is known as "Method of Least Squares"?
  Hint: Unit square is the basic unit in a graph.