Introduction to Machine Learning

Mentors:

- *Anupama Kolsur*
- *Rishik TS*
- *Shipra Mathur*
- *Surabhi*

Why "Learn"?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About"Learning"

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

  People who bought "Da Vinci Code" also bought "The Five People You Meet in Heaven"

- Build a model that is a good and useful approximation to the data.

What is Machine Learning?

- Machine Learning

  Study of algorithms that improve their performance at some task with experience
- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

Growth of Machine Learning

The machine learning market expected to grow from USD 1.03 Billion in 2016 to USD 8.81 Billion by 2022, Machine learning enabled solutions are being significantly adopted by organizations worldwide to enhance customer experience

Applications:

- Speech recognition, Natural language processing
- Computer vision
- Medical outcomes analysis
- Robot control
- Computational biology

Machine Learning Methods

- Supervised Learning
  - Regression
  - Classification
- Unsupervised Learning
- Reinforcement Learning

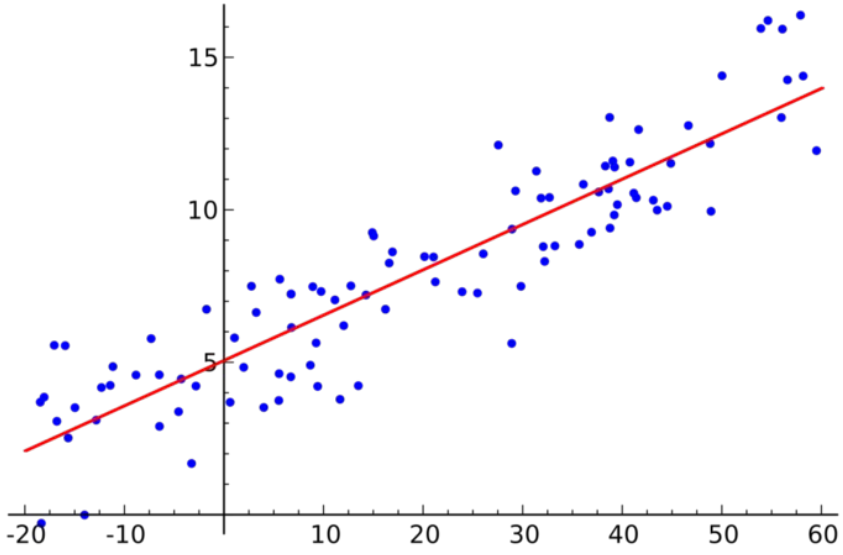
Supervised Learning

- Type of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output.
- The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).
- 2 types of supervised learning:

  Regression and classification
- eg. text classification problems

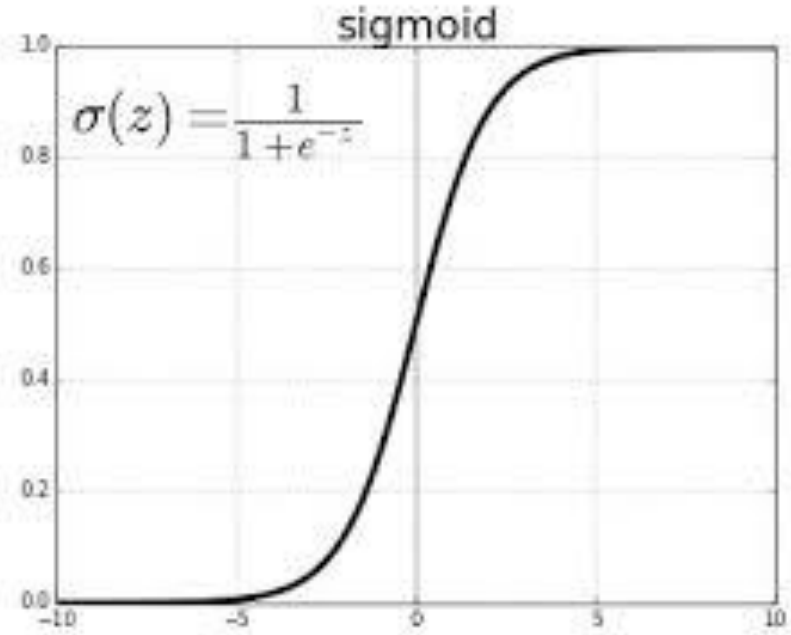
Linear Regression

- A supervised machine learning algorithm where the predicted output is continuous and has a constant slope.
- It's used to predict values within a continuous range
- Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable.



Logistic Regression

It is a classification algorithm in machine learning that uses one or more independent variables to determine an outcome. The outcome will take on only a small number of discrete values. We will mostly focus on binary classification problems.



Random Forest

- Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.
- It can be used for both Classification and Regression problems in ML.
- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Unsupervised Learning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Other applications: Summarization, Association Analysis
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

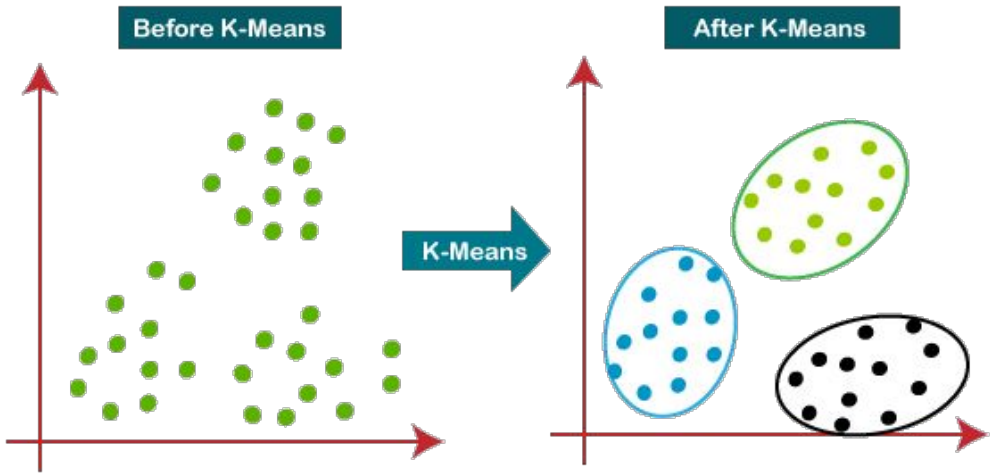
K Means Clustering

- Is a method of clustering basically which clusters n observations in k clusters.
- Each cluster will have a centroid such that square of euclidean distances of each point to its centroid is minimum.
- Applications:

    1) Sharding

    2) Image compression

    Many more...



Reinforcement Learning

- The model employs trial and error to come up with a solution to the problem. Based on the action performed, the model gets either rewards or penalties. Its goal is to maximize the total reward.
- Applications:
  - Game playing
  - Robot in a maze
  - Multiple agents, partial observability, ...