ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

UNIT IV

Learning Objectives

Machine Learning:

1. Unsupervised Learning
2. Dimensionality Reduction
3. Principal Component Analysis
4. K-Means Clustering
5. Ensemble Learning
6. Boosting and Bagging
7. Neural Networks
8. Types of Neural Networks
9. Activation Functions
10. Feed Forward, Back Propagation Algorithm
11. Recommender Systems, Content Based Recommendation

Unsupervised Learning

- Regression and classification problems belongs to a class of problems called supervised learning problems.
- When outcome or target variables are known i.e. when data is provided on the variable to be predicted, then the problem is a supervised learning problem.
- **Unsupervised learning** problems are those where the dataset **does not have any target or outcome variable**.

Have you heard about market basket analysis?

Unsupervised Learning (Contd.)

- In case of supervised learning algorithms we have patterns in our historical data and **define rules** saying that if **certain patterns** occur very often, then they can be used for making certain **recommendations**.
- This problem formulation is termed as **Association rule mining** and is an supervised learning problem.

Dimension Reduction

- Dimensionality reduction is transformation of data from **high dimensional space** to **low dimensional space**, retaining its meaningful properties of the original data.
- Complexity of the model increases with number of features.
- Popular technique to reduce dimensions:
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)

Dimension Reduction (Contd.)

- **High-dimensional data** can also lead to **overfitting**, where the model fits the training data too closely and does not generalize well to new data.
- There are two main approaches to dimensionality reduction:
  - feature selection and
  - feature extraction.

Feature Selection

- Feature selection involves selecting a subset of the original features that are most relevant to the problem at hand.
- The goal is to reduce the dimensionality of the dataset while retaining the most important features.
- There are several methods for feature selection:
  - Filter Methods
  - Wrapper Method
  - Embedded Method

Feature Extraction

- Feature extraction involves creating new features by combining or transforming the original features.
- There are several methods for feature extraction, including principal component analysis (PCA), linear discriminant analysis (LDA).
- Handling the high-dimensional data is very difficult in practice, commonly known as the *curse of dimensionality*. If the dimensionality of the input dataset increases, any machine learning algorithm and model becomes more complex.

Benefits of applying Dimensionality Reduction

- By reducing the dimensions of the features, the space required to store the dataset also gets reduced.
- **Less Computation training time** is required for reduced dimensions of features.
- Reduced dimensions of features of the dataset help in **visualizing the data quickly**.
- It removes the redundant features (if present) by taking care of **multicollinearity.**

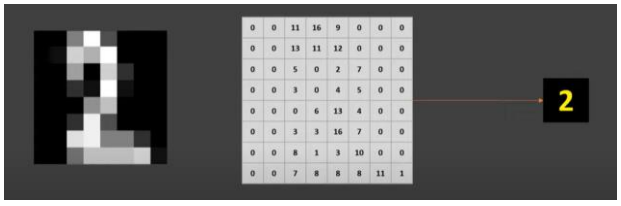
Disadvantages Dimensionality Reduction

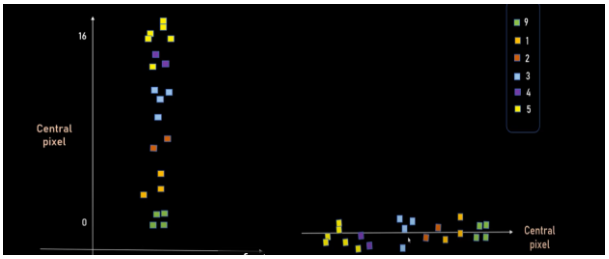
- Some **data may be lost** due to dimensionality reduction.
- In the PCA dimensionality reduction technique, sometimes the principal components required to consider are unknown.
- Interpreting transformed characteristics might be challenging.
- The independent variables become harder to comprehend as a result.

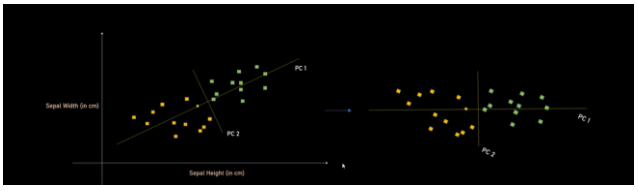
Principal Component Analysis

town	area	bathroom	plot	trees nearby	price
monroe	2600	2	8500	2	550000
monroe	3000	3	9200	2	565000
monroe	3200	3	8750	2	610000
monroe	3600	4	10200	2	680000
monroe	4000	4	15000	2	725000
west windsor	2600	2	7000	2	585000
west windsor	2800	3	9000	2	615000

Principal Component Analysis (Contd.)



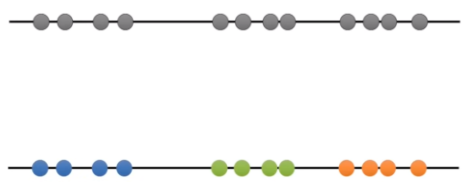

16
Central pixel
0
Central pixel
9
1
2
3
4
5


PC 1
Sepal Width (in cm)
PC 2
Sepal Height (in cm)
PC 1
PC 2

K- Means Clustering

- K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.
- To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids.
- It halts creating and optimizing clusters when either:
  - The centroids have stabilized — there is no change in their values because the clustering has been successful.
  - The defined number of iterations has been achieved.

K- Means Clustering (Contd.)

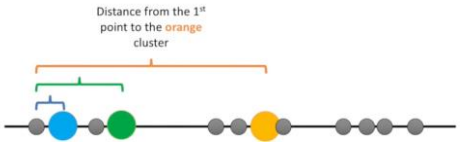


K- Means Clustering (Contd.)

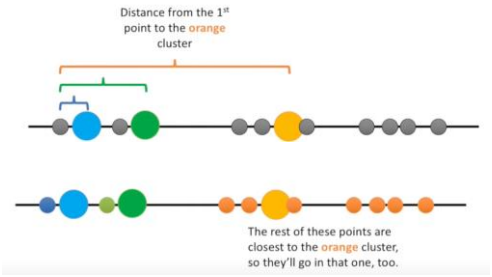
- Select K=3. It means we want to identify three clusters.
- Randomly select 3 data points.
- Measure the distance between first point and three initial clusters.

K- Means Clustering (Contd.)

- Select K=3. It means we want to identify three clusters.
- Randomly select 3 data points.
- Measure the distance between first point and three initial clusters.

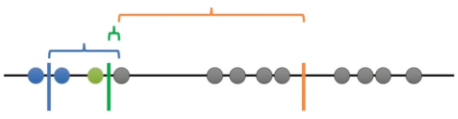


K- Means Clustering (Contd.)

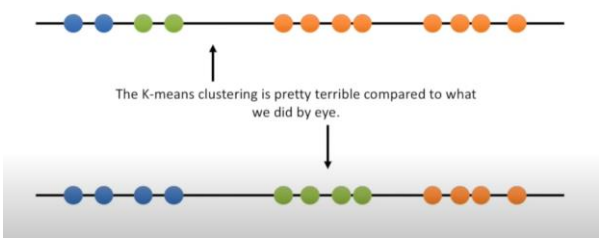


K- Means Clustering (Contd.)

- Then calculate the mean of each cluster just formed.

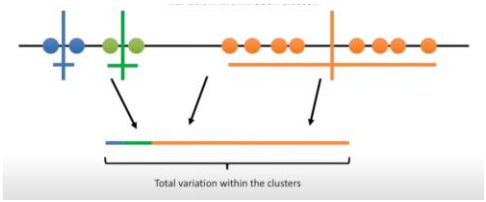


K- Means Clustering (Contd.)



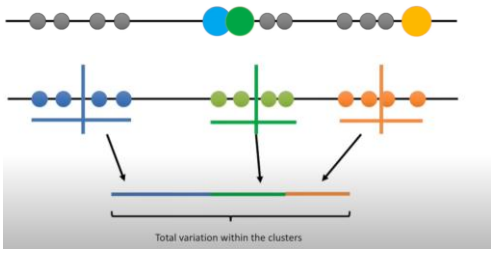
K- Means Clustering (Contd.)

- Now we can access the quality of the clustering by adding up the variation within each cluster.

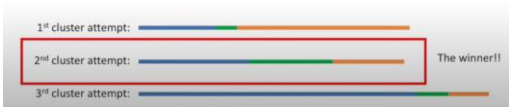


K- Means Clustering (Contd.)

- Now we can access the quality of the clustering by adding up the variation within each cluster.
-

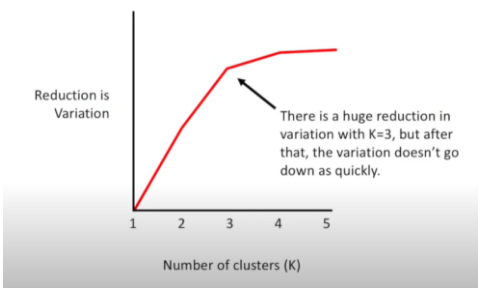


K- Means Clustering (Contd.)



- How to decide what value of K is good for clustering.
- Try all the values for K and quantify with the total variation.
- Each time K is increased the variation becomes smaller.

K- Means Clustering (Contd.)

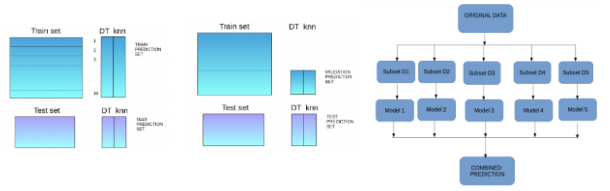


Ensemble Learning

- The ensemble methods in machine learning combine the insights obtained from multiple learning models to facilitate accurate and improved decisions.
- The underlying concept behind ensemble learning is to combine the outputs of diverse models to create a more precise prediction.
- Simple Ensemble Technique:
  - Max Voting
  - Averaging
  - Weighted Averaging

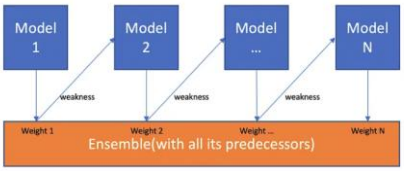
Ensemble Learning

- Advanced Ensemble Techniques:
  - Stacking
  - Blending
  - Bagging



Boosting

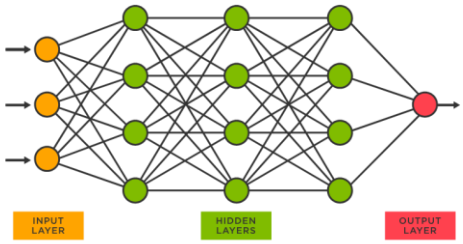
- Boosting is a sequential process, where each subsequent model attempts to correct the errors of the previous model. The succeeding models are dependent on the previous model.



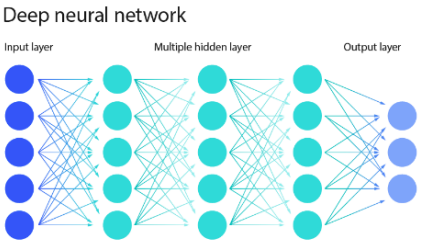
Neural Networks

- Artificial neural networks (ANNs) are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer.
- Each node, or artificial neuron, connects to another and has an associated weight and threshold.
- Neural networks rely on training data to learn and improve their accuracy over time.
- However, once these learning algorithms are fine-tuned for accuracy, they are powerful tools in computer science and artificial intelligence, allowing us to classify and cluster data at a high velocity.

Neural Networks



Neural Networks



Activation Function

- The activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The purpose of the activation function is to introduce non-linearity into the output of a neuron.

  Why do we need Non-linear activation function?
- A neural network without an activation function is essentially just a linear regression model. The activation function does the non-linear transformation to the input making it capable to learn and perform more complex tasks.
- Different types of activation functions:
  - Linear function
  - Sigmoid Function
  - Tanh Function
  - RELU function

Weighted input

Input	Desired Output
0	0
1	2
2	4

Now the output of your model when 'W' value is 3:

Input	Desired Output	Model output (W=3)
0	0	0
1	2	3
2	4	6

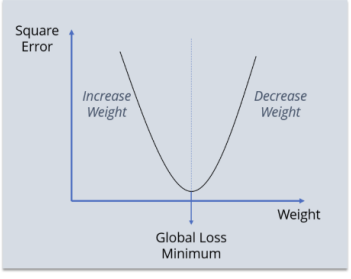
Input	Desired Output	Model output (W=3)	Absolute Error	Square Error
0	0	0	0	0
1	2	3	1	1
2	4	6	2	4

Weighted input

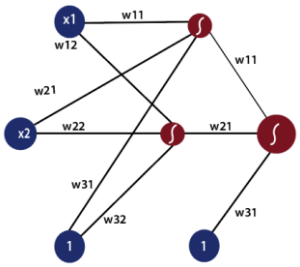
Input	Desired Output	Model output (W=3)	Absolute Error	Square Error	Model output (W=4)	Square Error
0	0	0	0	0	0	0
1	2	3	1	1	4	4
2	4	6	2	4	8	16

Input	Desired Output	Model output (W=3)	Absolute Error	Square Error	Model output (W=2)	Square Error
0	0	0	0	0	0	0
1	2	3	2	4	3	0
2	4	6	2	4	4	0

Minimum Loss



Feedforward Neural Network



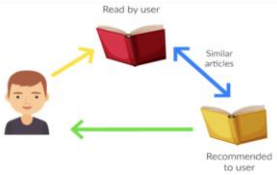
Back Propagation Neural Network

- Backpropagation is an algorithm that backpropagates the errors from the output nodes to the input nodes. Therefore, it is simply referred to as the backward propagation of errors.
- It computes the gradient of the loss function with respect to the network weights.
- The backpropagation algorithm works by computing the gradient of the loss function with respect to each weight via the chain rule, computing the gradient layer by layer, and iterating backward from the last layer to avoid redundant computation of intermediate terms in the chain rule.

https://www.edureka.co/blog/backpropagation/

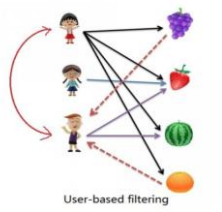
Recommender Systems

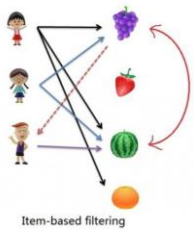
- A recommendation system is a subclass of Information filtering Systems that seeks to predict the rating or the preference a user might give to an item.
- It is an algorithm that suggests relevant items to users. Eg: In the case of Netflix which movie to watch, In the case of e-commerce which product to buy, or In the case of kindle which book to read, etc.



Types of Recommender Systems

- Content based filtering
- Collaborative Based Filtering
  - User-Based Collaborative Filtering
  - Item-Based Collaborative Filtering





Types of Recommender Systems (Contd.)

If No. of items is greater than No. of users go with user-based collaborative filtering as it will reduce the computation power and If No. of users is greater than No. of items go with item-based collaborative filtering