



Dr. Vishwanath Karad

MIT WORLD PEACE UNIVERSITY | PUNE

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

ASSIGNMENT / TEST BOOKLET

Student's Name : Vishwa Singh

Class : AML2 Division : Panel A Roll No.: PA10 Academic Year : 2020-2021

Subject : Advances in Machine Learning Assignment / Test No. : 2 Date : 23/12/2021

PLEDGE

I solemnly affirm that I have written this Assignment/Test based on my own preparation. I have neither copied it from others nor given it to others for coping. I know that this is to be submitted as a part of my submission at the end of the term.

Signature of the student

Q. No.	1	2	3	4	5	6	7	8	9	10	Total
Marks/Grade											

Name & sign of the faculty Member

(Please start writing assignment/ test from here)

Question 1 Explain fuzzy c means clustering

⇒ Fuzzy c means clustering algorithm which is very similar to the k-means algorithm and aims to minimize the objective function

$$\sum_{j=1}^k \sum_{x_i \in C_j} u_{ij}^m (x_i - u_j)^2$$

Where :

- u_{ij} is the degree to which an observation x_i belongs to a cluster c_j
- u_j is the center of the cluster j
- u_{ij}^m is the fuzzifier

The variable u_{ij}^m is

$$u_{ij}^m = \frac{1}{\sum_{l=1}^n \left(\frac{|x_i - c_l|}{|x_i - c_j|} \right)^{2/m-1}}$$

The parameter m is a real number greater than 1 ($1 \leq m < \infty$) and it defines the level of cluster fuzziness.

The fuzzy factors in between this is a soft clustering technique where each element has a probability of belonging to each cluster which depends on m .

Question 2

Hierarchical Clustering

K-means

- 1) Form dendrogram using top-down (divisive) or bottom-up (agglomerative) approach

Reposition centroids in order to minimize WCSS.

- 2) Relatively slower

Relatively faster

- 3) ~~Does~~ Returns a tree

Returns exactly k clusters



Question 3

Outlier Analysis: It is the process that involves identifying the anomalous observations in the dataset.

Outliers can be of two types:

- Univariate outliers: It can be found while looking at a distribution of values in a single feature space.
- Multivariate outliers: They can be found in n-dimensional space.

Detection method:

1) Z-Score: It is a metric or score of an observation that indicates how many standard deviations a data point is from the sample's mean, assuming a gaussian distribution.

2) DBSCAN (Density Based Spatial Clustering): In machine learning and data analytic clustering methods are useful tools that

understand the data better,

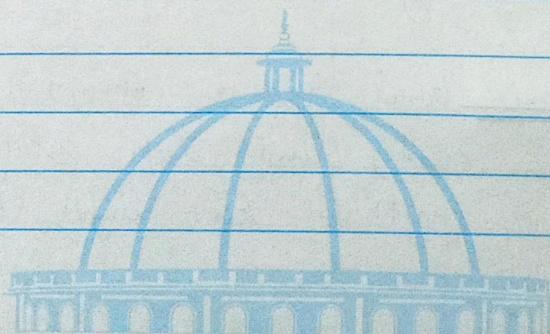
Decision defines different class points:

- Core points
- Border points
- Outliers

* Isolation Forest: The basic principle is that outliers are few and far from the rest of the observation. The algorithm randomly picks a feature from the space and a randomly split rule varying between the max & min.

Other methods:

- Probabilistic & statistical
- Linear regression models
- Information theory models



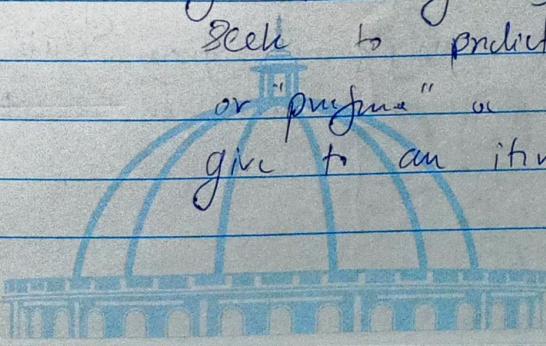
Question 4

⇒ Big Data : It is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of the traditional data management tools can store or process it efficiently.

Characteristics of big data :

- 1) Volume : The size of the data
- 2) Variety : Heterogeneous sources & the nature of data
- 3) Velocity : Speed of generation
- 4) Variability : Refers to the inconsistency which can be shown by the data.

Recommendation System : It is a subclass of information filtering system that seeks to predict the "rating" or "preference" of user would give to an item.



Methods:

- Content based filtering
- Collaborative filtering
- Hybrid filtering

Applications

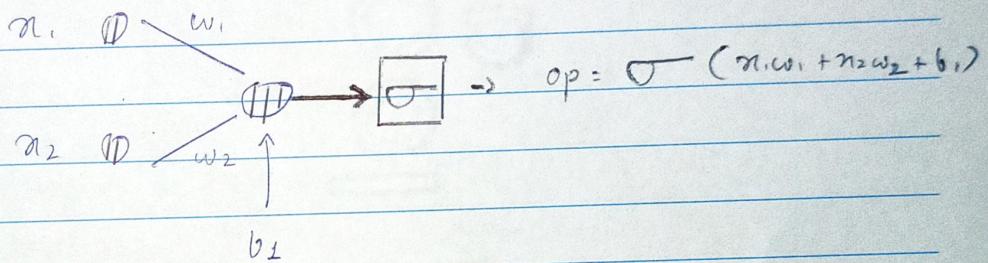
- 1) e-commerce: To recommend customer similar products
- 2) Media: To suggest some articles or posts that a user may like
- 3) Telecom: To suggest appropriate plans to customers

Lesson 5

Deep Learning: It is a subset of machine learning in AI that uses network capable of learning unsupervised from data that is unstructured or unlabeled

Thus require the use of an architecture called
deep neural network.

Perceptron :



Similarly, we can stack multiple nodes in a layer and multiple layers can be stacked together to create a DNN (Deep Neural Net).

Types of neural net:

- 1) Fully Connected : When every layer node of one layer is connected to every node of the next
- 2) Convolution : When only regions are connected in steps
- 3) Recurrent : Output of layer are connected to the input.

