Q1.	Choose the correct option for following questions. All the Questions are compulsory and carry equal marks
1.	K means clustering is based on _____ model of machine learning.
Option A:	Geometric model
Option B:	Probabilistic model
Option C:	Logical model
Option D:	Tree model
2.	The aim of _____ is to reduce the number of features in a dataset by generating new ones from the existing ones
Option A:	Feature selection
Option B:	Feature thresholding
Option C:	Feature extraction
Option D:	Feature cancellation
3.	The performance of classification is assessed using _____
Option A:	Square Matrix
Option B:	Confusion Matrix
Option C:	Diagonal Matrix
Option D:	Identity Matrix
4.	_____ is a harmonic mean of precision and recall.
Option A:	Specificity
Option B:	F1-Score
Option C:	Accuracy
Option D:	Sensitivity
5.	_____ measure is used for node splitting in Decision tree.
Option A:	Gini Index
Option B:	Mini Index
Option C:	Rand Index
Option D:	Maximum Index
6.	The _____ shows the trade-off between sensitivity (or TPR) and specificity (1 – FPR).
Option A:	POC Curve
Option B:	ROC Curve
Option C:	MOC Curve
Option D:	TOC Curve
7.	Difficulty of learning the joint probability in case of Bayes Classifier is solved by _____
Option A:	Simple Linear regression
Option B:	Logistic regression
Option C:	Naïve Bayes

Option D:	Multiple Linear regression
8.	_____ is stopping criteria in K Means Clustering
Option A:	re-assignments of data points to different clusters
Option B:	no re-assignments of data points to different clusters
Option C:	maximum decrease in the *sum of squared error*
Option D:	maximum change of centroids
9.	EM algorithm stands for_____
Option A:	Expectation-Maximisation
Option B:	Energy-Maximization
Option C:	Expectation-Minimisation
Option D:	Energy-Minimization
10.	Soft SVM is used when the data is_____
Option A:	clean
Option B:	Linear
Option C:	circular
Option D:	noisy

Q2. **(20 Marks Each)**	**Solve any Two Questions out of Three 10 marks each**
A	What is Machine Learning? Explain the issues in Machine Learning.
B	How the performance of classification and regression is assessed? Explain.
C	What is supervised learning? Explain the steps in developing Decision Tree algorithm.
Q3. **(20 Marks Each)**	**Solve any Two Questions out of Three 10 marks each**
A	Write a note on PCA.
B	What do you mean by ROC curve? Describe ranking and probability estimation trees.
C	Describe the application of Machine Learning in healthcare.
Q4.	**Solve any Two Questions out of Three 10 marks each**

	(20 Marks Each)	

A	For the following data, to construct the decision tree calculate Gini indexes a determine which attribute is root attribute.

Sr. No	Age	Income	Student	Credit_Rating	Byus_Computer
1	<=30	high	No	Fair	No
2	<=30	high	No	Excellent	No
3	31...40	high	No	Fair	Yes
4	>40	medium	No	Fair	Yes
5	>40	low	Yes	Fair	Yes
6	>40	low	Yes	Excellent	No
7	31...40	low	Yes	Excellent	Yes
8	<=30	medium	No	Fair	No
9	<=30	low	Yes	Fair	Yes
10	>40	medium	Yes	Fair	Yes
11	<=30	medium	Yes	Excellent	Yes
12	31...40	medium	No	Excellent	Yes
13	31...40	high	Yes	Fair	Yes
14	>40	medium	No	Excellent	No

B	The pairwise distance between 6 points is given below Using complete linkage proximity function in hierarchical clustering find and draw the resulting dendrogram.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

C	A spam filtering system has a probability of 0.95 to classify correctly a mail as spam and 0.10 probability of giving false positives. It is estimated that 0.5% of the mails are actual spam mails. Find the probability that, given a mail classified as spam by the system, the mail actually being spam