BE – VII/3(CBS(R)

25662

COMPUTER ENGINEERING   COURSE NO. PCS – 703

(Machine Learning)

Time Allowed: 3 Hours   Maximum Marks: 100

Note: Attempt five questions in all selecting at least two questions

from each Section. Each question carries 20 marks.

SECTION – A

Q1. (a) What is hypothesis in machine learning? Using suitable example define the   (10)
terms:
i) Hypothesis class ii) Version space
iii) Specific hypothesis iv) General hypothesis

(b) Compare and contrast Classification and regression. Identify two applications   (10)
of classification and regression.

Q2 (a) Differentiate between Supervised and Unsupervised learning. Explain with the   (10)
help of example.

(b) Show the final result of hierarchal clustering with complete link by drawing a   (10)
dendrogram.

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

Q3   Explain the Support Vector Machine (SVM) from the perspective of a non-   (20)
linear kernel by means of an algorithm. Derive the Margin of the support
vectors with an example and depict it with necessary diagrams.

Q4(a) Calculate precision, recall, F1 score and accuracy for the following cases in   (10)
dataset :
True Positive= 30, True Negative = 20, False Positive = 10, False Negative =
15   [Turn Over

(b) Use K-Means clustering to cluster the following data into two groups. Assume   (10)
cluster centroid are m1=2 and m2=4. The distance function used is Euclidean
distance. {2, 4, 10, 12, 3. 20, 30, 11, 25}

SECTION – B

Q5 (a) The following data set contains factors that determine whether tennis is played   (10)
or not. Using Naïve Bayes classifier, find the play prediction for the day
<Sunny, Cool, High, Strong>

DAY	OUTLOOK	TEMP	HUMIDITY	WIND	PLAY
Day1	Sunny	Hot	High	Weak	NO
Day 2	Sunny	Hot	High	Strong	NO
Day 3	Overcast	Hot	High	Weak	YES
Day 4	Rain	Mild	High	Weak	YES
Day 5	Rain	Cool	Normal	Weak	YES
Day 6	Rain	Cool	Normal	Strong	NO
Day 7	Overcast	Cool	Normal	Strong	YES
Day 8	Sunny	Mild	High	Weak	NO
Day 9	Sunny	Cool	Normal	Weak	YES
Day 10	Rain	Mild	Normal	Weak	YES
Day 11	Sunny	Mild	Normal	Strong	YES
Day 12	Overcast	Mild	High	Strong	YES
Day 13	Overcast	Hot	Normal	Weak	YES
Day 14	Rain	Mild	High	Strong	NO

(b) Discuss maximum likelihood hypothesis for predicting probabilities in   (10)
Bayesian learning.

Q6 (a) Describe the approaches for handling the overfitting of decision trees.   (10)
(b) Explain the basic decision tree learning algorithm.   (10)

Q7 (a) Consider a football game between two rival teams: Team 0 and Team 1.   (10)
Suppose Team 0 wins 95% of the time and Team 1 wins the remaining
matches. Among the games won by team 0, only 30% of them come from
playing on teams 1's football field. On the otherhand, 75% of the victories for
team 1 are obtained while playing at home. If team 1 is to host the next match
between the two teams, which team will most likely emerge as the winner?

(b) Explain the importance of the terms : (i) Hidden layer (ii) Generalization (iii)   (10)
Overfitting (iv) Stopping criterion

Q8 (a) Discuss the learning tasks and Q learning in the context of reinforcement   (10)
learning.
(b) What is Random Forest (RF)? Discuss reasons that improve RF based   (10)
classification compared to ID3.

------------------<>------------------