

Group – D

6. (a) Derive the linear regression formula for multiple dependent variables.
 (b) Explain briefly the difference between Input space and feature space. Also explain how the derived linear regression formula can be used for nonlinear cases.
7. (a) What are the various components of machine learning?
 (b) Consider the following statement -"Classification can be considered as a special case of regression" – Justify in favour or against of this statement.

6 + 6 =12**6 + 6 =12****Group – E**

8. Derive the Lagrangian for the optimization problem as defined by linear SVM – separable case. Derive the dual for the SVM.
9. Write Short Notes on **any three** of the following topics:
 (a) Single Perceptron Model
 (b) DbScan
 (c) Growth function
 (d) Back propagation Algorithm
 (e) Bias variance Tradeoff.

12**3 × 4 = 12****MACHINE LEARNING
(CSEN 5131)****Time Allotted : 3 hrs****Full Marks : 70***Figures out of the right margin indicate full marks.**Candidates are required to answer Group A and **any 5 (five)** from Group B to E, taking **at least one** from each group.**Candidates are required to give answer in their own words as far as practicable.***Group – A
(Multiple Choice Type Questions)**

1. Choose the correct alternative for the following: **10 × 1 = 10**
- (i) The goal in Naive Bayes classifier is to predict class label using,
 (a) posterior probability (b) prior probability
 (c) likelihood (d) evidence.
- (ii) In PCA transform the input data to a new data space _____.
 (a) of same dimension (b) of higher dimension
 (c) of smaller dimension (d) is none of (a), (b), (c).
- (iii) Decision trees are appropriate for the problems where,
 (a) Attributes are both numeric and nominal
 (b) Target function takes on a discrete number of values
 (c) Data may have errors
 (d) All of the above.
- (iv) Which of the following can act as possible termination conditions in *k*-means clustering algorithm?
 1. a fixed number of iterations.
 2. assignment of observations to clusters does not change between iterations, except for cases with a bad local minimum.
 3. centroids do not change between successive iterations.
 (a) 1 and 2 (b) 1 and 3
 (c) 2 and 3 (d) 1,2 and 3

- (v) Which of the following gives non-linearity to a neural network.
 (a) Convolution operator (weighted sum of the inputs)
 (b) Stochastic gradient descent
 (c) Sigmoid activation
 (d) Non-zero bias.
- (vi) When a model performs well on training data (the data on which the algorithm was trained) but does not perform well on test data (new or unseen data), we say that the model is
 (a) Overfitting
 (b) Generalizing
 (c) Regularizing
 (d) None of the above.
- (vii) After SVM learning, each Lagrange multiplier α_i takes either zero or non-zero value. What does it indicate in each situation?
 (a) A non-zero α_i indicates the data point i is a support vector, meaning it touches the margin boundary.
 (b) A non-zero α_i indicates that the learning has not yet converged to a global minimum.
 (c) A zero α_i indicates that the data point i has become a support vector data point, on the margin.
 (d) A zero α_i indicates that the learning process has identified support for vector i .
- (viii) The back-propagation algorithm learns a globally optimal neural network with hidden layers.
 (a) Always True
 (b) Always False
 (c) Mostly true
 (d) Mostly False.
- (ix) Consider the following two statements:
 1. "Noise objects are always outliers."
 2. Outliers are always noise objects."
 (a) Both are true
 (b) (1) is true and (2) is false
 (c) (1) is false and (2) is true
 (d) Both are false.
- (x) A hard margin in SVM indicates:
 (a) that the SVM allows very low error in classification
 (b) that the SVM allows high amount of error in classification
 (c) that the SVM guarantees no over-fitting of training data set
 (d) None of the above.

Group – B

2. (a) Why naïve Bayesian classification is called naïve?
 (b) Use Naive Bayes' classifier to predict whether a person defined by the tuple (age = youth, income = medium, student = yes, credit_rating = fair) buys a computer or not. The training data is as follows:

RID	Age	Income	Student	Credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	Yes
14	senior	medium	no	excellent	no

2 + 10 = 12

3. (a) Explain the single layer neural network (perceptron) for classification using the backpropagation algorithm for a data set D consisting of the training tuples and their associated target values.
 (b) Describe what is likely to happen when a learning rate is used that is too large, and when one is used that is too small. How can one optimize the learning rate?
 (c) Explain the purpose of the momentum term that is often included in the Back-Propagation learning algorithm.

8 + 2 + 2 = 12**Group – C**

4. (a) What are the benefits of dimension reduction of input data in the context of clustering problem?
 (b) Explain how Principal component analysis can be used for dimension reduction using an appropriate example.
5. (a) How to choose a reasonably good of "k" before clustering a given dataset using k-means clustering technique.
 (b) What are the limitations of k-means clustering technique?
 (c) What are the basic differences between ordinary and kernel k-means clustering technique?

4 + 8 = 12**4 + 4 + 4 = 12**