MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

Paper Code : PCCAIML601 Machine Learning Applications

UPID : 006916

Time Allotted : 3 Hours *Full Marks :70*

The Figures in the margin indicate full marks.
Candidate are required to give their answers in their own words as far as practicable

Group-A (Very Short Answer Type Question)

1. Answer *any ten* of the following : [1 x 10 = 10]

(I) In what type of regularization, weight sparsity is obtained?

(II) Given 3 data points in 2d space, (1,1), (2,2) and (3,3). What is the first principal component?

(III) Variance - Covariance matrix is real and symmetric. State True or False.

(IV) How many layers can we add in logistic regression model other than input and output?

(V) Can we model non-linear relationships of data with a linear regression?

(VI) As uncertainty increases entropy decreases. Is it True or False?

(VII) In logistic regression model, we predict the data in continuous form. Is it True or False?

(VIII) Boosting trees are built sequentially while random forest build trees in parallel. Is it True or False?

(IX) We have 1% training error and 2% validation error in case of low bias and low variance. Is it True or false?

(X) The values of Lagrange multipliers corresponds to the support vectors can be less than zero. Is it True or false?

(XI) Reducing the number of leaves in decision tree will increase the variance. Is it True or False?

(XII) Llyod's algorithm is always guaranteed to converge. State True or False.

Group-B (Short Answer Type Question)

Answer *any three* of the following : [5 x 3 = 15]

2. Can we get nonlinear decision boundary in decision tree? How axis labels are used in decision tree? What is pruning? What are stumps? [5]

3. How many local minimum can be encountered while solving the optimization for maximizing margin for SVM? Justify your answer. [5]

4. Eigenface approach is one practical application of PCA. Let us assume that we have 500 face images having dimension 64x64. How could you construct the covariance matrix to find the principal components? [5]

5. What is elbow method? State the application of it. What type of loss do we use in performing elbow method? [5]

6. Say h= $g(b+w_1x_1+w_2x_2,)$ and $w_1=1, w_2=1, b=-3$. Write down the equation for decision boundary. Assume that we have four classes. How can we represent these four classes using one hot encoding? [5]

Group-C (Long Answer Type Question)

Answer *any three* of the following : [15 x 3 = 45]

7. (a) Why are SVMs fast? State 2-3 advantages of it. [3]

(b) Why are SVMs more accurate than logistic regression? State 2-3 reasons. [3]

(c) Find the width of the gutter in SVM, or find the distance from support vectors to hyperplane. [6]

(d) What is kernel trick? How a nonlinearly separable data can be linearly separable using kernel? [3]

8. (a) In a project we have a set of data from patients who have visited ABCD hospital during the year 2022. A set of features (e.g., bmr, weight) have been also extracted for each patient. The goal of the project is to decide whether a new visiting patient has any of diabetes, heart disease, or neural disease (a patient can have one or more of these diseases). What are the features here? What is the basic course of problem? Regression or Classification? [3]

(b) We have decided to use a neural network to solve this problem. We have two choices: either to train a separate neural network for each of the diseases or to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer [6]

(c) In our project we first ask our model to predict whether a patient has a disease, and if the classifier is 80% confident that the patient has a disease, then we will collect additional patient features like temperatures etc. In this case, which classification methods do you recommend: neural networks, decision tree, or naive Bayes? Justify your answer. [6]

9. (a) Say, we have 5 data points A,B,C,D,E. and we are allowed to use 6 weak classifiers. We use the decision tree stumps. [1]
The datapoints are at (1,5,A),(5,5,B),(3,3,C),(1,1,D),(5,1,E) in the form (x_1,x_2, data point).
Find out the misclassification when $x_1>2$, $x_1<2$, $x_1>4$, $x_1<4$, $x_1>6$, $x_1<6$.

(b) Update the weights for each data point up to round 3. [6]

(c) In each round, calculate the error rate. [3]

(d) In each round, calculate the strong classifier. [3]

(e) Can we stop after round 3? Explain. [2]

10. (a) Can we use SGD for both ridge regression and lasso regression? If yes, explain why. If not, explain why and what other techniques we can use for optimization. [2]

(b) Derive the closed form solution for ridge regression model considering maximum likelihood estimation. [5]

(c) What is weight decay in ridge regression? [1]

(d) Why data preprocessing is a must in linear regression.? [1]

(e) Let we have the data X={(1.2, 3043), (2.4,506.3), (3.5,1127)}. Derive the prepossessed data. [2]

(f) If weights are initialized as (0,0) and learning rate = 3, for the above data find the value of weights after 2nd epoch. Consider regularizing parameter value of 10. [4]

11. (a) Consider the following design matrix. X=[[4,1],[2,3],[5,4],[1,0]], 4 2D data points. We want to reduce the dimension from 2 to 1 and represent the data. Show the detail process with numerical steps. [5]

(b) Consider 4 datapoints as (1,0), (4,1), (2,3) and (5,4). We want one dimensional representation of data. Draw the principal components with direction. Show all the projections of all four sample points onto the principal direction. [6]

(c) What is the disadvantage in k-means clustering? How this can be overcome by K means+ ? [4]

*** END OF PAPER ***