B.Tech. III 5th Semester CS365 Machine Learning

Institute Elective Subject - Mid Semester Examination September 2021 - 1 October 2021 @ 11:30 am

u19cs012@coed.svnit.ac.in Switch account

Draft saved

Your email will be recorded when you submit this form

* Required

Consider a binary classification problem. Suppose, a model is trained on a linearly separable training set, and now I get a new labelled data point which is correctly classified by the model, and far away from the decision boundary. If I now add this new point to my earlier training set and re-train, in which cases is the learnt decision boundary likely to change? * 1 point

- ○ When a model is a perceptron.
- ○ When a model is an SVM.
- ◉ When a model is Gaussian discriminant analysis.
- ○ None of these

Suppose a SVM is trained on n > 100 data points in 2-dimensional plane, yielding a hyperplane with exactly 2 support vectors. If one more data point is added and the classifier is retained, what is maximum possible number of support vectors for the new hyperplane assuming all data points are linearly separable? * 1 point

- ○ 2
- ○ 3
- ○ n
- ◉ n+1

Neural networks emulate which types of complex functions with many parameters * 1 point

- ◉ Linear
- ○ Nonlinear
- ○ Discrete
- ○ Continuous

The Indian Railways have been experimenting a machine learning method which attempts to predict whether a train will arrive at its final destination on time or not, using a number of input features corresponding to weather conditions, train priorities, ongoing repair works etc. (for this purpose, `on time' is defined as no more than 10 minutes after its scheduled time). The methods have been tested on a common set of 500 train runs, and the results are as follows: 1. the number of trains predicted on time: (actually on time:131, actually late:155) 2. the number of trains predicted late: (actually on time:19, actually late:195). A simple probabilistic model is set up. Select the correct tuple from the following. The tuple consists of three values (the prior probability of a train being late, probability of a late prediction if the train is on time called false positive rate (FPR), probability of a late prediction if the train is in fact late called true positive rate (TPR)). Values are calculated with two digits precision. * 3 points

- (0.7,0.56, 0.44)
- (0.3, 0.87, 0.44)
- (0.7, 0.13, 0.56)
- (0.3,0.87, 0.13)

Backpropagation algorithm performs * 1 point

- Pattern mapping
- Prediction
- Function approximation
- All of these

In neural networks, nonlinear activation functions such as sigmoid and tanh * 1 point

- ( ) are applied only to the output units
- ( ) always output values between 0 and 1
- (•) help to learn nonlinear decision boundaries
- ( ) speedup the gradient calculation in backpropagation

The kernel trick * 1 point

- ( ) can be applied to every classification algorithm
- ( ) is commonly used for dimensionality reduction
- (•) exploits the fact that in many learning algorithms, the weights can be written as a linear combination of input points
- ( ) for regression it can solve a d x d linear system instead of an n x n system given n sample points with d features

What is a type of hyperplane in geometrical model of support vector machine? * 1 point

- (•) a plane with 1 dimensional fewer than number of input attributes
- ( ) a plane with 2 dimensional fewer than number of input attributes
- ( ) a plane with 1 dimensional more than number of input attributes
- ( ) a plane with 2 dimensional more than number of input attributes

Bayes rule basically answers any query by using * 1 point

- ○ Full distribution
- ◉ Joint distribution
- ○ Marginal distribution
- ○ All of these

The input to the credit scoring system consists of records of the customers having 3 attributes (Occupation, Age, Salary in rupees) and each record is labelled as (Defaulter, Genuine). The records consist of 1. Industrial, 39, 34k, defaulter 2. Industrial, 22, 40k, defaulter 3. Professional, 30, 27k, genuine 4. Professional, 27, 33k, defaulter 5. Professional, 40, 20k, genuine 6. Industrial 27, 30k, genuine 7. Professional, 50, 70k, defaulter 8. Industrial, 33, 26k, genuine 9. Industrial, 30, 45k, defaulter 10. Professional, 45, 27k, genuine. Select the correct tuple from the following for the set of the following queries:(i) An average age of defaulter (ii) Apriori class probability (iii) Standard deviation of salary of defaulter (iv) Probability of a person is defaulter given it is industrial (v) Probability of a person is genuine given it is professional. Values are calculated with two digits precision. Answer tuples are * 4 points

- ○ (35,0.5,45.66 ,0.4 ,0.4)
- ○ (33.6,0.5,13.51 ,0.6 ,0.4)
- ○ (35,0.5,45.66 ,0.4 ,0.6)
- ◉ (33.6,0.5,13.51 ,0.6, 0.6)

Bayes rule can be used for * 1 point

- ○ solving complex queries
- ○ decreasing complexity
- ◉ answering probabilistic query
- ○ None of these

A neural network has a neuron model with input x with weight w that goes into neuron. Say in the network, Neuron A has input IA that has WA. Neuron C inputs the output of neuron A with weight WAC. Neuron C also has input IC that has weight WC. Neurons output the sum of the products of each input with their respective weight. What is the output of neuron C if IA = 0.5, WA = 2, WAC = 1, IC = 1, and WC = 3? * 1 point

- ○ 3
- ○ 4
- ◉ 5
- ○ 6

Disadvantages of Naïve Bayes Classifier * 1 point

- ◉ Naïve Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between
- ○ It performs well in Multi-class predictions as compared to the other
- ○ Naïve Bayes is one of the fast and easy ML algorithms to predict a class of
- ○ It is the most popular choice for text classification problems

Machine learning is a subset of * 1 point

- ○ Deep learning
- ◉ Artificial Intelligence
- ○ Data learning
- ○ None of these

Application of machine learning methods to large databases is called * 1 point

- ◉ data mining
- ○ artificial intelligence
- ○ big data computing
- ○ internet of things

In what type of learning labelled training data is used * 1 point

- ○ unsupervised learning
- ◉ supervised learning
- ○ reinforcement learning
- ○ active learning

Data used to build a data mining mode * 1 point

- (●) training data
- ( ) validation data
- ( ) test data
- ( ) hidden data

Which algorithm is preferred given a set of reviews of few Netflix series marked as positive, negative and neutral, the new series are to be marked? * 1 point

- (●) supervised learning
- ( ) unsupervised learning
- ( ) semi-supervised learning
- ( ) reinforcement learning

Machine learning technique that helps in detecting the outlier in data * 1 point

- ( ) Clustering
- ( ) Classification
- (●) Anamoly detection
- ( ) All of these

Which of the following example can be addressed using supervised learning Algorithm? * 1 point

- (●) given email labelled as spam or not spam, learn a spam filter
- ( ) given a set of news articles found on the web, group them into set of articles about the same story
- ( ) given a database of customer data, automatically discover market segments and group customers into different market segments
- ( ) find the patterns in market basket analysis

It is assumed that if the grade for internal examination is high in a class, the grade for external examination will also be high. The random samples of 15 students in the class are presented as a pair of (internal marks, external marks)=(X,Y): (15, 49), (23, 63), (18, 58), (23, 60), (24, 58), (22, 61), (22, 60), (19, 63), (19, 60), (16, 52), (24, 62), (11,30), (24, 59), (16, 49), (23, 68). A regression model Y = (a + b X) + w. w is an error. Evaluate the parameter values of a and b using least square approach that is minimize the sum of square of errors. w error is the difference between the predicted and actual value of Y. (Hint: b and a can be found using minimizing the objective function that is the sum of square of errors. b comes out to be (covariance(X,Y)/variance(X)) and a is (average(Y) – b average(X)). Values are calculated with two digits precision. Select the correct tuple (a,b): * 5 points

- ( ) (19.93, 5.68)
- (●) (19.05, 1.89)
- ( ) (18.70, 6.72)
- ( ) (18.42, 4.78)

Submit Clear form

Google Forms