Indian Institute of Technology, Bombay

GNR 652 – Machine Learning for Remote Sensing-I

Tutorial Quiz-I

Question 1. The nodes with homogeneous class distribution are preferred in decision trees has

   a. High degree impurity

   b. Low degree impurity

   c. Can't say

   d. None of the above

Question 2. Let us suppose there are a total of 10 samples for two-class classification. The GINI index for
i.) 4 samples in class C1 and 6 samples in class C2
ii.) 7 samples in class C1 and 3 samples in class C2 are

   a. 0.33 and 0.50 respectively

   b. 0.33 and 0.48 respectively

   c. 0.48 and 0.42 respectively

   d. 0.42 and 0.48 respectively

Question 3. In a linear regression model, the test error decreases with an increase in training data and the training error increases. But the training is quite lower than the testing error. The main reason behind this behavior is

   a. High variance

   b. High model bias

   c. High estimation bias

   d. Low variance

Question 4. The gradient of a continuous and differentiable function (**multiple answers can be correct**):

   a. is zero at a minimum

   b. is non-zero at a maximum

   c. is zero at a saddle point

   d. decreases as you get closer to the minimum

Question 5. Insufficient model capacity, under-fitting, and over-fitting leads to

   a. High bias, high bias, and high variance respectively

   b. Low bias, high bias, and high variance respectively

   c. High bias, low, bias, and low variance respectively

   d. Low bias, low bias, and low variance respectively

Question 6. The computational complexity of Gradient descent is (**N denotes the number of data points and D denotes the dimensionality**):

   a. Linear in D

   b. Linear in N

   c. Polynomial in D

   d. Dependent on the number of iterations

Question 7. Which of the following assumptions do we make while deriving linear regression parameters?
   1. The true relationship between dependent y and predictor x is linear.
   2. The model errors are statistically independent.
   3. The errors are normally distributed with a 0 mean and constant standard deviation.
   4. The predictor x is non-stochastic and is measured error-free.

   a. 1, 2, and 3

   b. 1, 3, and 4

   c. 1 and 3

   d. All of the above

Question 8. Suppose I applied a logistic regression model on data and got training accuracy X and testing accuracy Y. Now I want to add a few new features in data. Select option(s) which are correct in such cases.
Note: Consider the remaining parameters are the same.
   1. Training accuracy always decreases.
   2. Training accuracy always increases or remains the same.
   3.Testing accuracy always decreases.
   4.The predictor x is non-stochastic and is measured error-free.

   a. Only 2

   b. Only 1

   c. Only 3

   d. Only 4

Question 9. Suppose we fit "Lasso Regression" to a data set, which has 100 features $(X_1, X_2 \ldots X_{100})$. Now, we re-scale one of these features by multiplying with 10 (say that feature is X_1), and then refit Lasso regression with the same regularization parameter.

   a. It is more likely for X_1 to be excluded from the model

   b. It is more likely for X_1 to be included in the model

   c. Can't say

   d. None of the above

Question 10. Which of the following functions can be used as an activation function in the output layer if we wish to predict the probabilities of n classes $(p_1, p_2 .. p_k)$ such that sum of p over all n equals to 1?

   a. Softmax

   b. ReLU

   c. Sigmoid

   d. Tanh