The Analytics Edge

Fall 2018

Concept Questions

1. Suppose we use linear regression on a training dataset and get a sum of squared error equal to e_1. Using this model, we obtain a sum of squared error in the test set equal to e_2. We add some more new features (predictor variables) in the dataset and refit the model. Select the best option.

   (a) The training error e_1 always decreases or remains the same.

   (b) The training error e_1 always increases or remains the same.

   (c) The test error e_2 always decreases or remains the same.

   (d) The test error e_2 always increases or remains the same.

2. The statement, "the p-value is 0.001" is equivalent to the statement that "there is a 0.1% probability that the null hypothesis is true". True or False?

3. If you get a p-value of 0.1, it means that when the null hypothesis is true, a value of the test statistic as or more extreme than what was observed occurs in about 10% of all samples. True or False?

4. Suppose we solve a linear regression problem and obtain the optimal estimates $\hat{\beta}_0, \ldots, \hat{\beta}_p$. Show that the average value of the residuals with these optimal estimates will be always 0.

5. Suppose a 95% confidence interval for the slope β of a linear regression of y on x is given by $-3.5 < \beta < -0.5$. Then a two- sided test of the hypothesis $H_0 : \beta = -1$ would result in rejection of H_0 at the 1% level of significance.