Worksheet assignment 3

Machine learning

Q1 – C

Q2 – A

Q3 – B

Q4 – B

Q5 – C

Q6 – B

Q7 – A

Q8 – A

Q9 – B&D

Q10 – A & C

Q11 – A, B & D

Q12 – A

Q13 – A & B

Q14 - Explain Linear Regression?

Ans - Linear regression is a statistical method used to model the linear relationship between a dependent variable and one or more independent variables. It is a supervised learning algorithm used to predict a continuous outcome (dependent variable) based on one or more predictor variables (independent variables).

The basic idea of linear regression is to find the best fitting line through the data points that represents the relationship between the independent and dependent variables. The line is defined by a set of coefficients that are estimated from the training data. Once the coefficients are estimated, the model can be used to make predictions on new data.

In linear regression, the goal is to minimize the difference between the predicted values and the actual values of the dependent variable. This is done by minimizing the sum of the squared errors (SSE) between the predicted and actual values. The coefficients of the model are estimated using a method called ordinary least squares (OLS) that minimizes the SSE.

Linear regression has several assumptions, including linearity, homoscedasticity, independence, and normality of errors. Violations of these assumptions may result in biased or unreliable estimates.

Q15 What is difference between simple linear and multiple linear regression?

Ans - The main difference between simple linear regression and multiple linear regression is the number of independent variables used to predict the dependent variable.

In simple linear regression, there is only one independent variable (predictor variable) used to predict the dependent variable. The relationship between the independent and dependent variables is represented by a straight line. For example, in a simple linear regression model to predict housing prices, the independent variable could be the size of the house, and the dependent variable could be the price of the house.

In multiple linear regression, there are two or more independent variables used to predict the dependent variable. The relationship between the independent and dependent variables is represented by a multi-dimensional hyperplane. For example, in a multiple linear regression model to predict housing prices, the independent variables could be the size of the house, the number of bedrooms, the age of the house, and the location of the house, and the dependent variable could still be the price of the house.

Worksheet assignment 3 - Python

Q1 – None

Q2 – C

Q3 – B

Q4 – A

Q5 – C

Q6 – D

Q7 – B

Q8 – A

Q9 – A, B, C & D

Q10 – A & C

Worksheet Assignment 3 – Statistics

Q1 – c

Q2 – b

Q3 – a

Q4 – a

Q5 – c

Q6 – c

Q7 – a

Q8 – b

Q9 – b

Q10 – c

Q11 – d

Q12 – b

Q13 - How do you find the test statistic for two samples?

Ans - To find the test statistic for two samples, you would typically perform a two-sample hypothesis test using a t-test or z-test, depending on the situation. The exact method used will depend on the sample size, whether the population standard deviation is known or unknown, and whether the samples are independent or dependent.

Here are the general steps for finding the test statistic for two samples using a t-test:

1. State the null and alternative hypotheses. The null hypothesis typically assumes no difference between the means of the two samples, while the alternative hypothesis assumes a difference.
2. Determine the significance level and the appropriate test statistic. For a two-sample t-test, the test statistic is calculated as the difference between the sample means divided by the standard error of the difference.

3. Calculate the sample means, sample standard deviations, and sample sizes for each group.
4. Calculate the standard error of the difference between the two sample means.
5. Calculate the t-statistic using the formula:

   t = (x1 - x2) / (SE)

   where x1 and x2 are the sample means, and SE is the standard error of the difference.

6. Determine the degrees of freedom and use them to find the critical value of t from the t-distribution table or calculator.
7. Compare the calculated t-statistic with the critical value of t to determine whether to reject or fail to reject the null hypothesis.

Q14 - How do you find the sample mean difference?

Ans - To find the sample mean difference, you first need to have two sets of sample data, each with their own sample mean. The sample mean is the average of all the data points in the sample.

Once you have the two sample means, you can find the sample mean difference by subtracting one sample mean from the other. Mathematically, the formula for the sample mean difference is:

Sample mean difference = Sample mean 1 - Sample mean 2

For example, let's say you have two sets of data, and you want to find the sample mean difference between them:

Set 1: 3, 5, 7, 9, 11 (sample mean = 7) Set 2: 2, 4, 6, 8, 10 (sample mean = 6)

To find the sample mean difference between these two sets, you would subtract the sample mean of Set 2 from the sample mean of Set 1:

Sample mean difference = 7 - 6 = 1

So the sample mean difference between Set 1 and Set 2 is 1. This means that, on average, the data values in Set 1 are 1 unit higher than the data values in Set 2.Top of Form

Q15 - **What is a two sample t test example?**

Ans - A two sample t-test is used to compare the means of two independent groups. Here is an example scenario where a two sample t-test can be applied:

Suppose a researcher is interested in comparing the effectiveness of two different weight loss programs. She recruits two groups of participants, each containing 30 individuals. One group is assigned to follow Program A, while the other group is assigned to follow Program B. At the end of the 12-week program, the researcher measures the weight loss for each participant and calculates the mean weight loss for each group. The data is summarized as follows:

Program A: Mean weight loss = 8 pounds, Standard deviation = 2 pounds Program B: Mean weight loss = 6 pounds, Standard deviation = 3 pounds

The researcher wants to determine if there is a significant difference in the mean weight loss between the two programs. To do this, she performs a two-sample t-test. The null hypothesis is that there is no difference in the mean weight loss between the two programs, while the alternative hypothesis is that there is a difference.

The researcher calculates the test statistic, which is the t-score, using the following formula:

t = (mean of Group A - mean of Group B) / (pooled standard deviation * sqrt(1/n1 + 1/n2))

where n1 and n2 are the sample sizes of each group and the pooled standard deviation is calculated as follows:

pooled standard deviation = sqrt(((n1 - 1) * s1^2 + (n2 - 1) * s2^2) / (n1 + n2 - 2))

s1 and s2 are the standard deviations of Group A and Group B, respectively.

Assuming a significance level of 0.05, the researcher compares the calculated t-score to the critical t-value from the t-distribution with (n1 + n2 - 2) degrees of freedom. If the calculated t-score is greater than the critical t-value, the researcher rejects the null hypothesis and concludes that there is a significant difference in mean weight loss between the two programs.