Quiz 10 - Regression, Cluster Analysis and Association Analysis

1. What is the main difference between classification and regression?

- In classification, you're predicting a number, and in regression, you're predicting a category.
- There is no difference since you're predicting a numeric value from the input variables in both tasks.
- **In classification, you're predicting a category, and in regression, you're predicting a number.**
- In classification, you're predicting a categorical variable, and in regression, you're predicting a nominal variable.

2. Which of the following is NOT an example of regression?

- Predicting the price of a stock
- Estimating the amount of rain
- **Determining whether power usage will rise or fall**
- Predicting the demand for a product

3. In linear regression, the least squares method is used to

- Determine the distance between two pairs of samples.
- Determine whether the target is categorical or numerical.
- **Determine the regression line that best fits the samples.**
- Determine how to partition the data into training and test sets.

4. How does simple linear regression differ from multiple linear regression?

- In simple linear regression, the input has only categorical variables. In multiple linear regression, the input can be a mix of categorical and numerical variables.
- **In simple linear regression, the input has only one variable. In multiple linear regression, the input has more than one variables.**

- In simple linear regression, the input has only categorical variables. In multiple linear regression, the input has only numerical variables.
- They are the just different terms for linear regression with one input variable.

5. The goal of cluster analysis is

- To segment data so that differences between samples in the same cluster are maximized and differences between samples of different clusters are minimized.
- To segment data so that all samples are evenly divided among the clusters.
- To segment data so that all categorical variables are in one cluster, and all numerical variables are in another cluster.
- **To segment data so that differences between samples in the same cluster are minimized and differences between samples of different clusters are maximized.**

6. Cluster results can be used to

- Determine anomalous samples
- Segment the data into groups so that each group can be analyzed further
- Classify new samples
- Create labeled samples for a classification task
- **All of these choices are valid uses of the resulting clusters.**

7. A cluster centroid is

- The mean of all the samples in the two closest clusters.
- **The mean of all the samples in the cluster**
- The mean of all the samples in the two farthest clusters.
- The mean of all the samples in all clusters

8. The main steps in the k-means clustering algorithm are

- **Assign each sample to the closest centroid, then calculate the new centroid.**
- Calculate the centroids, then determine the appropriate stopping criterion depending on the number of centroids.
- Calculate the distances between the cluster centroids, then find the two closest centroids.
- Count the number of samples, then determine the initial centroids.

9. The goal of association analysis is

- To find the most complex rules to explain associations between as many items as possible in the data.
- To find the number of outliers in the data
- **To find rules to capture associations between items or events**
- To find the number of clusters for cluster analysis

10. In association analysis, an item set is

- A transaction or set of items that occur together
- A set of transactions that occur a certain number of times in the data
- A set of items that two rules have in common
- A set of items that infrequently occur together

11. The support of an item set

- **Captures the frequency of that item set**
- Captures how many times that item set is used in a rule
- Captures the number of items in that item set
- Captures the correlation between the items in that item set

12. Rule confidence is used to

- Identify frequent item sets
- Determine the rule with the most items
- Measure the intuitiveness of a rule
- **Prune rules by eliminating rules with low confidence**