Group Coaching Session - IX

The session will be conducted after the students have completed all the modules in basic machine learning, i.e., linear, logistic, and clustering.

Objective

Since you are nearing the end of course 2 and have learnt about three fundamental machine learning algorithms this session will be firstly focused mostly on clearing your doubts surrounding machine learning and business problem-solving. Then, based on the assessment of your group over the past few weeks, the instructor will summarise the important learnings from course 2 machine learning algorithms and provide a walkthrough of some applications and/or interview questions from the topics on the same.

Agenda

1. Part-I: Doubt Resolution + P2P (30 mins)
2. Part-II: Focused Teaching - Course 2 ML Concepts Summary & Applications/Interview Questions (60 mins)

Questions

Please find a broad pool of questions below. This pool might contain two kinds of questions:

- Most frequent doubts that our learners face in these topics
- Common interview questions from these topics

Linear Regression

1. What are the assumptions of linear regression?
2. Assume that you have made predictions using Linear Regression, when would you be sure that the predictions made can be acceptable?
3. What do you understand by R2-score? When are r-squared and adjusted r-squared the same?
4. Does a weak Pearson's correlation coefficient always imply that the variables are not correlated?
5. What are some common drawbacks of the linear regression model? What are some ways to overcome them?
6. What is multicollinearity? How do you identify and deal with multicollinearity in your dataset?
7. How does RFE work?
8. Explain the working of gradient descent.
9. What is heteroscedasticity? What are the consequences, and how can you overcome it?
10. How do you interpret a linear regression model?
11. How is hypothesis testing used in linear regression?

Logistic Regression

1. Why can't linear regression be used instead of logistic regression for binary classification?
2. What is the likelihood function?
3. What are odds and log odds?
4. Why is logistic regression widely preferred in the industry?
5. What is the maximum likelihood estimator or MLE?
6. Why is accuracy not a good measure for evaluating classification problems?
7. What are some other evaluation metrics that overcome the disadvantages of accuracy?
8. Explain the use of a ROC curve.
9. How do you decide an optimal cutoff point for logistic regression?
10. Explain the different elements of a confusion matrix.

Clustering

1. Explain the steps of K-means Clustering algorithm. Mention the key steps that need to be followed and how the algorithm works.
2. Explain the types of segmentation that can be considered while solving a business problem using clustering.
3. What are the different proximity functions or distance metrics used for the K-means algorithm?
4. What are the issues with random initialization of centroids in K-means algorithm and how to overcome it?
5. How are outliers handled by the K-means algorithm?
6. What is the objective function for measuring the quality of clustering in case of the K-means algorithm with Euclidean distance?
7. Is K-means clustering suitable for all shapes and sizes of clusters?
8. What are the types of hierarchical clustering?
9. What are the disadvantages of agglomerative hierarchical clustering?
10. Is validation required for clustering? If yes, then why is it required?