Daily Machine Learning Interview Questions

26. What is Cross-Validation?

Cross-Validation in Machine Learning is a statistical resampling technique that uses different parts of the dataset to train and test a machine learning algorithm on different iterations. The aim of cross-validation is to test the model's ability to predict a new set of data that was not used to train the model.

Cross-validations avoids them overfitting of data. K-Fold Cross Validation is the most popular resampling technique that divides the whole dataset into K sets of equal sizes.

27. What are the different methods to split a tree in a decision tree algorithm?

- **Variance:** Splitting the nodes of a decision tree using the variance is done when the target variable is continuous.

- **Information Gain:** Splitting the nodes of a decision tree using Information Gain is preferred when the target variable is categorical.

- **Gini Impurity:** Splitting the nodes of a decision tree using Gini Impurity is followed when the target variable is categorical

28. How does the Support Vector Machine algorithm handle self-learning?

The SVM algorithm has a learning rate and expansion rate which takes care of self-learning.

The learning rate compensates or penalizes the hyperplanes for making all the incorrect moves while the expansion rate handles finding the maximum separation area between different classes.

29. What is the difference between Lasso and Ridge regression?

Lasso(also known as L1) and Ridge(also known as L2) regression are two popular regularization techniques that are used to avoid overfitting of data. These methods are used to penalize the coefficients to find the optimum solution and reduce complexity.

The Lasso regression works by penalizing the sum of the absolute values of the coefficients. In Ridge or L2 regression, the penalty function is determined by the sum of the squares of the coefficients.

30. What are the assumptions you need to take before starting with linear regression?

There are primarily 5 assumptions for a Linear Regression model:

- Multivariate normality
- No auto-correlation
- Homoscedasticity
- Linear relationship
- No or little multicollinearity

Thank You