**Regression**:

**1. Difference between training and test data**

Training dataset is used to train the model and test dataset is used to evaluate the model.

**2. Evaluation metrics of regression model**

Mean Absolute error

Mean Squared error

R squared score

**3. Type of data unsuitable for regression.**

Categorical data is unsuitable for regression

**4. Mean Absolute error vs Mean Squared error**

Mean Absolute error measures the difference between actual value and predicted value whereas mean squared error measures the squared difference between actual and predicted values

**5. Coefficients of a regression model.**

The coefficients of a regression model is based on a straight line equation Y=mX+C

In this, m is slope/weight and it varies with each input

C is intercept/bias and it is constant for the model

If the coefficient is positive, then the input variable will be directly proportional to the output variable. If the input increases, the output will also increase.

If the coefficient is negative, then the input variable will be inversely proportional to the output variable. If the input increases, the output will decrease.

**6. Residual in regression**

The difference between the actual value and the predicted value is the residual.

If the residual is more, the model is of poor performance

If the residual is less, the model is of good performance.

**7. Why cross validation is important in regression?**

Cross validation is important because it is used to create the model with the best tuning parameter and also to compare the performance of the model by evaluating the model with different algorithms for the same dataset.

**8. R2 score – 0.85**

If the R squared value is 0.85, then the performance of the model is average.

**9. What steps can you take if your regression model is too slow to converge?**

Data pre-processing can be tried if the regression model is too slow to converge.

**10. Regression model is underperforming. What steps can you take?**

If the model is underperforming, then the performance can be improved by selecting a different algorithm or by changing the hypertuning parameters.

**Classification:**

1. **Classification in machine learning:**

Classification in Machine learning works on the dataset by categorising the dataset into groups (eg: Yes/No, patient diagnosed with disease or not) using different algorithms.

1. **Classification algorithms:**

* KNN – K Nearest Neighbor
* Naïve Bayes
* Logistic Regression
* Decision Tree Classifier
* Random Forest Classifier
* Support Vector Classifier

1. **Metrics used to evaluate Classification models:**

* Precision
* Recall
* F1-Score
* Accuracy
* Macro Average
* Weighted Average

1. **Confusion Matrix**

Confusion matrix is used to evaluate the performance of the model. The components are

* True positive
* False positive
* False negative
* True negative

1. **Classification vs Regression:**

Classification is used to categorise the dataset into groups whereas the regression works on the numerical dataset and used to predict the continuous output values based on the input variable.

1. **ROC** **Curve**

ROC – Receiver Operating Characteristic is used to evaluate the performance of the classification models.