Applied Data Science

CLA-1

AP20110010782

AP20110010787

AP20110010380

AP20110010718

AP20110010369

## **Q2 Explain the evaluation metrics of linear regression model.**

Evaluation metrics for a linear regression model.

Evaluation metrics are a measure of how good a model performs and how well it approximates the relationship. Let us look at **MSE, MAE, R-squared, RMSE and Adjusted R-squared.**

## **1)Mean Squared Error (MSE)**

The most common metric for regression tasks is MSE. It has a convex shape. It is the average of the squared difference between the predicted and actual value. Since it is differentiable and has a convex shape, it is easier to optimize. MSE penalizes large errors.

## **2)Mean Absolute Error (MAE)**

This is simply the average of the absolute difference between the target value and the value predicted by the model. Not preferred in cases where outliers are prominent. MAE does not penalize large errors.

## **3)R-squared**

This metric represents the part of the variance of the dependent variable explained by the independent variables of the model. It measures the strength of the relationship between your model and the dependent variable. If the data points are very close to the regression line, then the model accounts for a good amount of variance, thus resulting in a high R² value.

If R² is high (say 1), then the model represents the variance of the dependent variable.

If R² is very low, then the model does not represent the variance of the dependent variable and regression is no better than taking the mean value, i.e. you are not using any information from the other variables.

A Negative R² means you are doing worse than the mean value. It can have a negative value if the predictors do not explain the dependent variables at all

## **4)Root Mean Squared Error (RMSE)**

This is the square root of the average of the squared difference of the predicted and actual value.

R-squared error is better than RMSE. This is because R-squared is a relative measure while RMSE is an absolute measure of fit (highly dependent on the variables — not a normalized measure).

Basically, RMSE is just the root of the average of squared residuals. We know that residuals are a measure of how distant the points are from the regression line. Thus, RMSE measures the scatter of these residuals. RMSE penalizes large errors.

## **5)Adjusted R-squared**

The main difference between **adjusted R-squared** and R-square is that **R-squared** describes the amount of variance of the dependent variable represented by every single independent variable, while **adjusted R-squared** measures variation explained by only the independent variables that actually affect the dependent variable.

R² tends to increase with an increase in the number of independent variables. This could be misleading. Thus, the adjusted R-squared penalizes the model for adding furthermore independent variables (k in the equation) that do not fit the model.