Metrics – Machine Learning Regression

Diptangshu Banik

Overview

- Used to determine the performance of a model.
- Classification and Regression has different sets of metrics.
- There are no metrics for Clustering as there is no Y label in this type of problems.

- Classification metrics are based on how many predictions are correct or incorrect.

- Regression metrics are based on how close is the predicted value from the actual value.

Regression

- MSPE
- MSAE
- R Square
- Adjusted R Square

Classification

- Precision-Recall
- ROC-AUC
- Accuracy
- Log-Loss

Unsupervised Models

- Rand Index
- Mutual Information

Others

- CV Error
- Heuristic methods to find K
- BLEU Score (NLP)

Regression

Mean Absolute Deviation:

Mean Absolute Deviation is the average of the difference between the Original Values and the Predicted Values. It gives us the measure of how far the predictions were from the actual output. However, they don’t gives us any idea of the direction of the error i.e. whether we are under predicting the data or over predicting the data.

MACHINE LEARNING

- Mathematically,

$$MeanAbsoluteError = \frac{1}{N}\sum_{j=1}^{N}\left|y_j - \hat{y}_j\right|$$

Mean Absolute Error :

MAE is the average of the absolute difference between the predicted values and the observed values.

As we take square of the error, the effect of larger errors become more pronounced then smaller error, hence the model can now focus more on the larger errors.

It is a linear score. All individual differences are weighed equally in the average.

Mathematically,

$$MeanSquaredError = \frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2$$

Root Mean Square Error(RMSE Score)

It represents the sample standard deviation between the predicted values and the observed values.

RMSE is a popular formula to measure the error rate of a regression model. However, it can only be compared between models whose errors are measured in the same units.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

MACHINE LEARNING

Relative Squared Error

Unlike RMSE, the relative squared error (RSE) can be compared between models whose errors are measured in the different units.

$$RSE = \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{\sum_{i=1}^{n}(\bar{a} - a_i)^2}$$

a = actual target

p = predicted target

$\bar{a}$ =mean of actual values

MACHINE LEARNING

Relative Absolute Error

Like RSE , the relative absolute error (RAE) can be compared between models whose errors are measured in the different units.

$$RAE = \frac{\sum_{i=1}^{n} |p_i - a_i|}{\sum_{i=1}^{n} |\bar{a} - a_i|}$$

Coefficient of Determination (R^2)

The coefficient of determination (**R^2**) summarizes the explanatory power of the regression model and is computed from the sums-of-squares terms.

R^2 describes the proportion of variance of the dependent variable explained by the regression model. If the regression model is "perfect", SSE is zero, and R^2 is 1. If the regression model is a total failure, SSE is equal to SST, no variance is explained by regression, and R^2 is zero.

MACHINE LEARNING

$$\textit{Coefficient of Determination} \rightarrow \quad R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$\textit{Sum of Squares Total} \rightarrow \quad SST = \sum (y - \bar{y})^2$$

$$\textit{Sum of Squares Regression} \rightarrow \quad SSR = \sum (y' - \overline{y'})^2$$

$$\textit{Sum of Squares Error} \rightarrow \quad SSE = \sum (y - y')^2$$

$\bar{y}$ = the mean of actual values

y = the actual values

y' = the predicted values

$\overline{y'}$ = the mean of all predicted values