Performance Metrics

Monday, July 19, 2021 5:09 PM

Performance Metrics are used to measure the performance of Model / Function which is obtained by Training Data.

For Regression :

- Mean Absolute Error :

$$\Sigma_{i=1}^{n} \left| y_{actual} - Y_{predicted} \right| / n$$

- Mean Squared Error :

$$\Sigma_{i=1}^{n} \left(y_{actual} - Y_{predicted} \right)^2 / n$$

- Root Mean Squared Error : $\sqrt{MSE}$

- Median Absolute Deviation : Median $\left| y_{actual} - Y_{predicted} \right|$
  It is powerful metric but its having more time complexity.

- R2_score : Its not a very good metric to tell performance , it only tells about how good the model fits to data. It is also called Coefficient of Determination.

SS_{total} : **Sum of Squared Error.**

$$\Sigma_{i=1}^{n} \left(y_{act} - \overline{Y_{pred}} \right)^2$$

where $\overline{Y_{pred}}$ get from mean model of data where we just impute mean of y_{input} in data.

$\mathbf{SS_{residual}}$: Sum of Squared Residual

$$\Sigma_{i=1}^{n} \left(y_{act} - Y_{pred} \right)^2$$

where Y_{pred} get from regression model.

SS_{total} has larger Error than $\mathbf{SS_{residual}}$ because we just imputing mean in model.

So,

$$R^2 = 1 - \frac{SS_{residual}}{\mathbf{SS_{total}}} \qquad 0 \leq R^2 \leq 1$$

Case 1 :

If $SS_{residual} = 0$,

$$R^2 = 1 - \frac{0}{\mathbf{SS_{total}}} = 1$$ which means model performs really well

Case 2 :

If $SS_{residual} = SS_{total}$

$R^2 = 1 - 1 = 0$ model is performing like simple mean model.

Case 3 :

If $SS_{residual} > SS_{total}$

$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = -ve$

Case 4 :

If $SS_{residual} < SS_{total}$

$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = 0 - 1$

For Classification :

- Accuracy : $\frac{No\ of\ Correctly\ Classified\ points}{Total\ No\ of\ Points} * 100$

Why we need another metrics for Classification ?

:: Imbalance Dataset :- Very large difference in No. of different Class labels. Some are very large and other are small. Imbalance Dataset have high ratio of different class labels and in real life we get many times Imbalance Dataset.

So, accuracy cannot be considered a good metrics because we not enough data for all class labels.

- Confusion Matrix : Data having 60 +ve : 53+ve , 7 -ve
  40 -ve : 35-ve , 5 +ve

Predicted \ Actual	+ve	--ve
+ve	TP 53	FP Type II error 5
-ve	FN Type I error 7	TN 35

Accuracy = $\frac{53+35}{53+5+7+35}$

Diagonal values are correctly classified. It should be high as high possible. Model is performing well if diagonals are high.

True Positive Rate (TPR) : TP / P = 53/ 60 = 0.8833 Also called **Sensitivity**
True Negative Rate (TNR) : TN / N = 35/40 = 0.875 Also called **Specificity**

False Positive Rate (FPR) : FP / N = 5/40 = 0.125
False Negative Rate (FNR) : FN / P = 7/ 60 = 0.1167

TPR + FNR = 1
TNR + FPR = 1

TPR and TNR are high as possible.

Precision and Recall :

Precision : Of all points, the model predicts to be +ve, what percentage of points are actually +ve

Precision (0-1) = $\frac{TP}{TP+FP}$ = $\frac{53}{53+5}$ = 53/ 59 = 0.8983

Recall : Of all points, the model predicts to be +ve, what percentage of points are predicted to be +ve. Also called Sensitivity/ TPR

Recall (0-1) = $\frac{TP}{TP+FN}$ = $\frac{53}{53+7}$ = 53/60 = 0.8833

Good Model should have HIGH both Precision and Recall close to 1.

F1 - score :

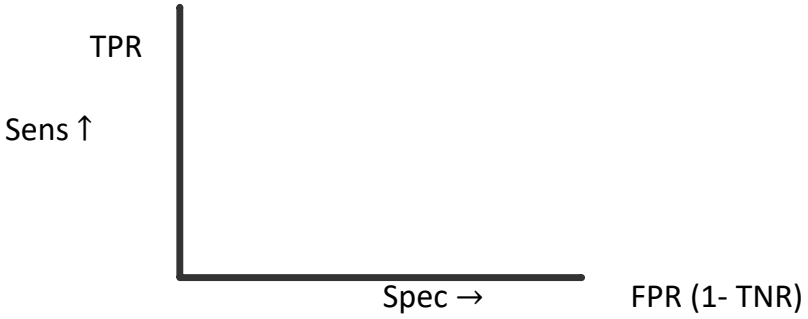
Harmonic Mean of Precision and Recall

= 2 * $\frac{Precision * Recall}{Precision+Recall}$

If one of them is low, then F1 score also be low.

ROC - AUC (Receiver Operation Characteristic Curve and Area Under Curve) :

Works only for Binary Classification. It affects from imbalance data.



Area of Graph = 1

Any model which have Area under Curve greater than Area of Diagonal = 0.5 , is considered to be good model.

Here M2 is performing good as compared to M1.
If the model has Area =1, It considered as best model.

Log Loss Function : The smaller the better.

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Note : Only Confusion Matrix, Precision and Recall , F1 score is considered to be good for Multiclass Classification.