Model Performance Metric

Weijia Xiong, Xinru Wang

7/08/2020

Classification Accuracy

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made}$$

It works well only if there are equal number of samples belonging to each class.

For example, consider that there are 98% samples of class A and 2% samples of class B in our training set. Then our model can easily get 98% training accuracy by simply predicting every training sample belonging to class A.

When the same model is tested on a test set with 60% samples of class A and 40% samples of class B, then the test accuracy would drop down to 60%. Classification Accuracy is great, but gives us the false sense of achieving high accuracy.

Confusion Matrix

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Confusion Matrix

There are 4 important terms:

- **True Positives**: The cases in which we predicted YES and the actual output was also YES.
- **True Negatives**: The cases in which we predicted NO and the actual output was NO.
- **False Positives**: The cases in which we predicted YES and the actual output was NO.
- **False Negatives**: The cases in which we predicted NO and the actual output was YES.

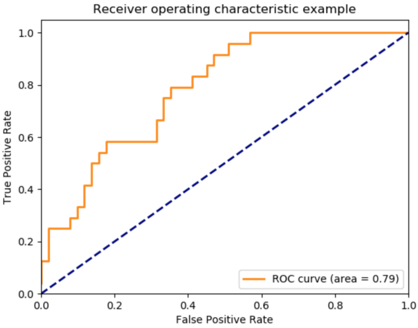
Accuracy for the matrix can be calculated by taking average of the values lying across the "main diagonal"

$$Accuracy = \frac{TruePositive \ + \ TrueNegative}{Total\ Sample}$$

Area Under Curve

- **True Positive Rate (Sensitivity)**: the proportion of positive data points that are correctly considered as positive.
- **True Negative Rate (Specificity)**: the proportion of negative data points that are correctly considered as negative.
- **False Positive Rate**: the proportion of negative data points that are mistakenly considered as positive.

FPR and TPR both are computed at varying threshold values such as (0.00, 0.02, 0.04, ..., 1.00) and a graph is drawn. AUC is the area under the curve of plot False Positive Rate vs True Positive Rate at different points in [0, 1].



F1 Score

- **Recall**: the proportion of positive data points that are correctly considered as positive. Equals to Sensitivity.
- **Precision**: the proportion of cases that considered as positive that are actually positive.

F1 Score is needed when you want to seek a balance between Precision and Recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Continuous Outcome

Mean Absolute Error

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

Mean Squared Error

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

The advantage of MSE being that it is easier to compute the gradient, whereas Mean Absolute Error requires complicated linear programming tools to compute the gradient.

Relative Square Error

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

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