

Food for Thought

What are some different methods for measuring the performance of a classification algorithm?

Which algorithms tend to perform better in terms of ROC AUC?

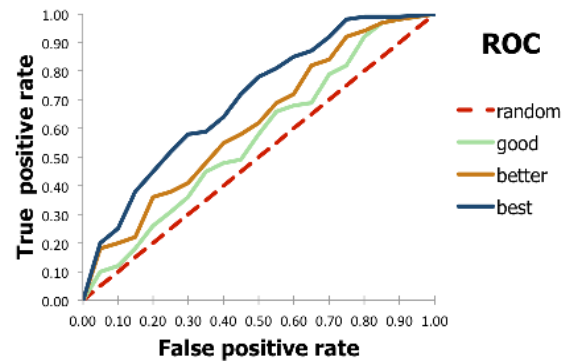
What predefined Python methods and classes exist for plotting ROC curves and calculating their AUC?

What is a ROC Curve?

A Receiver Operating Characteristic (ROC) Curve is a measurement of success rates based upon the threshold settings. When analyzing the accuracy of data, more data put in to classify and train will likely lead to more successful results.

What does a ROC Curve demonstrate?

It shows the tradeoff between sensitivity and specificity most importantly. To know how good the test is, look at the curve of the line. If the curve is close to the left border and the top border of the graph, the better the test is, and likewise the closer to a 45 degree line, the worse it is. It's also important to know that if the area under the curve (AUC) is closer to 1, the test is better; closer to 0, worse.



How do we measure different aspects of a ROC Curve?

There are multiple formulas associated with the ROC curve, and they're all fairly simple, using the values with which we calculated precision and recall.

$$\text{FPR} = 1 - \text{Specificity}$$

$$\text{TPR / Recall / Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

True Positive Rate measures the actual accuracy; Specificity is the complete negative, which measures the failures of a system. FPR, or False Positive Rate, measures the inaccuracy of the system.

ROC Curve Applications:

- It can assess the accuracy of diagnostic tests in clinical trials.
- It can assess the ability of an MRI to predict diseases.

Code

```
#Need y_train_5 and y_scores from elsewhere in the chapter to run this code
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score
fpr, tpr, thresholds = roc_curve(y_train_5, y_scores)
def plot_roc_curve(fpr, tpr, label=None):
    plt.plot(fpr, tpr, linewidth=2, label=label)
    plt.plot([0, 1], [0, 1], "k--")
    plt.axis([0, 1, 0, 1])
    plt.xlabel("False Positive Rate")
    plt.ylabel("True Positive Rate")
plot_roc_curve(fpr, tpr)
plt.show() #Unnecessary in Jupyter Notebook
roc_auc_score(y_train_5, y_scores)
```

Resources

<https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>

<http://gim.unmc.edu/dxtests/roc2.htm>