LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

Evaluation

Carlos J. Costa

Evauation

- After a data scientist has chosen a target variable and completed the prerequisites of transforming data and building a model, one of the final steps is evaluating the model's performance.

Confusion Matrix

- This matrix describes an output of “yes” vs. “no”.
- These two outcomes are the “classes” of each example.

actual		Predict: No	Predict: yes
	No	90	10
	yes	5	95

Confusion Matrix

- To better interpret the table, it is possible to see it in terms of:
  - true positives (TP): number of positive records rightly predicted as positive
  - true negatives (TN): number of negatives records rightly predicted as negative
  - false positives (FP): number of negative records wrongly predicted as positive
  - false negatives (FN): number of positive records wrongly predicted as negative.

actual		Predict: No	Predict: yes
	No	True Negative	False Positive
	yes	False Negative	True Positive

Accuracy

- Overall performance of the model
- Overall, how often is our model correct?

$$Accuracy=(TP+TN)/(TP+FP+TN+FN)$$

actual		Predict: No	Predict: yes
	No	True Negative	False Positive
	yes	False Negative	True Positive

Precision or positive predictive value (PPV)

- How accurate the positive predictions are
  - Precision=TP/(TP+FP)
- Precision helps when the costs of false positives are high.
  - e.g. detect skin cancer

actual	Predict: No	Predict: yes
No	True Negative	False Positive
yes	False Negative	True Positive

Recall or true positive rate (TPR)

- Coverage of actual positive sample

$$Recall=TP/(TP+FN)$$

- Recall helps when the cost of false negatives is high.
  - e.g. detect nuclear missil

actual \ Predict	No	yes
No	True Negative	False Positive
yes	False Negative	True Positive

Specificity or true negative rate (TNR)

- Coverage of Actual negative Sample

$$Specificity = TN/(TN+FP)$$

actual		Predict: No	Predict: yes
	No	True Negative	False Positive
	yes	False Negative	True Positive

F1 Score

- Hybrid metric useful for unbalanced samples

  F1=2((precision x recall)/(precision + recall))

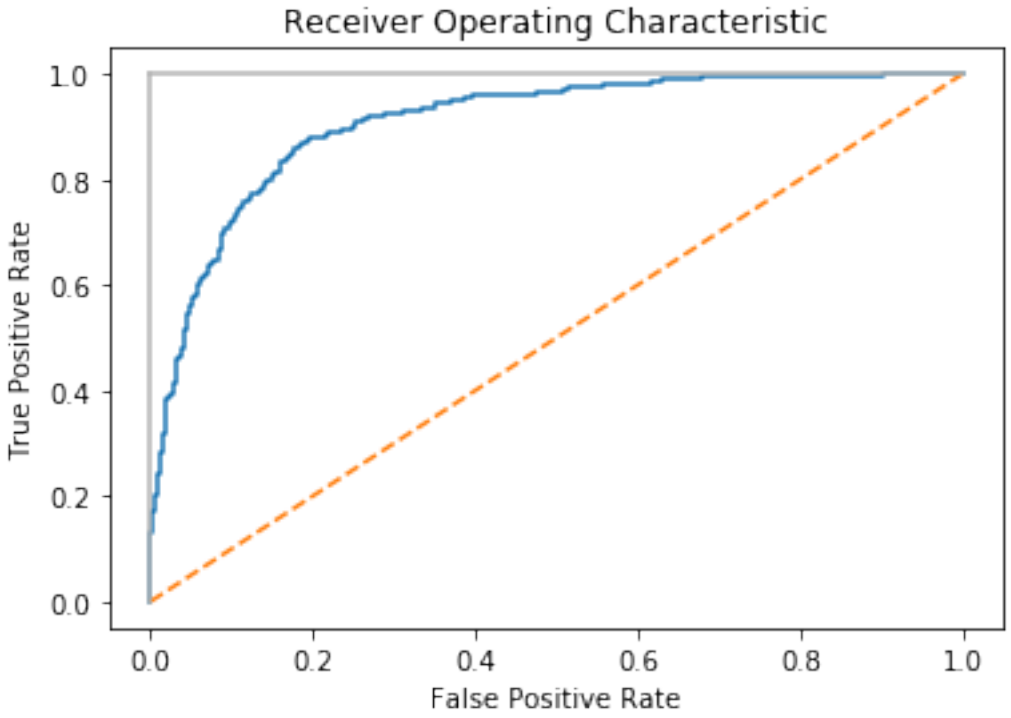
- a good F1 score means:
  - low false positives &
  - low false negatives
- correctly identifying real threats
- not disturbed by false alarms.

		Predict	
actual		No	yes
	No	True Negative	False Positive
	yes	False Negative	True Positive

ROC

- Receiver operating characteristic curve
- Specially useful in presence of binary non balanced datasets.
- ROC Charts present the balance between True Positive rate (recall) and False Positive rate in a graphical way,
- ROC Charts are available through the roc_curve method in the sklearn.metrics

ROC


Receiver Operating Characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0