***Obtaining an estimated roc – auc score.***

This Kaggle competition uses a method called “Receiver Operator Characteristic – Area Under Curve” to determine the precision and accuracy (or sensitivity and specificity) of the method used in prediction.

Scikit-Learn includes an roc and roc-auc method based on the true and predicted dependent variables, which is great. We have written a method called return\_roc\_auc() that reads in and applies a transform to the directory passed and then divides that data (either using a straight cross-validation, or you can use a 5-fold) and trains the specified classifier on some data and predict on the rest. Then we can see where the pipeline predicted correctly and where it didn’t.

To call the return\_roc\_auc() method you need to pass the name of the study, the directories that are to be read in, the data directory, the list of the Scans and the classifier to be tested:

import epilepsyTools as eT

import TransformsBH as transforms

import sys

from sklearn.ensemble import RandomForestClassifier

name = “StatsTransform\_RandomForest”

dataDirectory = “/storage/Epilepsy/”

listOfScans = ['Dog\_1', ' Dog\_2', ' Dog\_3', ' Dog\_4', ' Dog\_5', 'Patient\_1', 'Patient\_2']

transform = transforms.fft

classifier = RandomForestClassifier(n\_estimators=3000, min\_samples\_split=1, bootstrap=False, n\_jobs=4, random\_state=0)

score = eT.return\_roc\_auc(name, dataDirectory, listOfScans, transform, classifier)

Score will contain a real number of the calculated roc/auc score.

return\_roc\_auc has a number of optional arguments, which are described in the ImplementationNotes.py file.