GENERAL ASSEMBLY

INTRO TO DATA SCIENCE

EVALUATING CLASSIFICATION MODELS

AGENDA

- Miscellaneous SciKit-Learn Tips
- Review Logistic Regression and Naive Bayes
- Evaluating Classification Models: ROC Curve

MISCELLANEOUS SCIKIT LEARN TIPS

SKLEARN API: ESTIMATORS

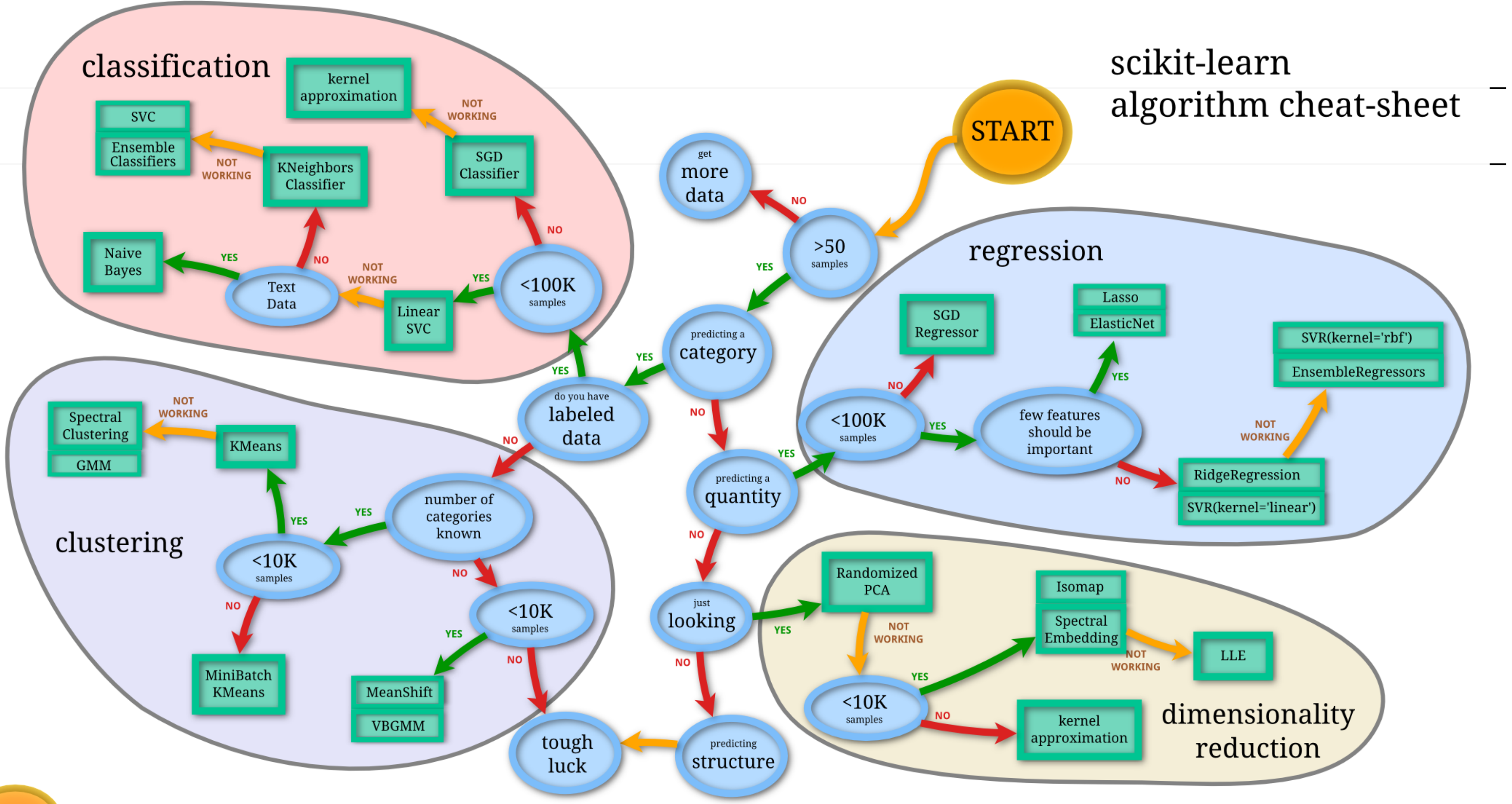
“An estimator is any object that learns from data; it may be a classification, regression or clustering algorithm...”

SKLEARN API: ESTIMATORS

```
class Estimator(object):

        def fit(self, X_train, y_train=None):
        """Fits estimator to data. """
        # set state of ``self``

    def predict(self, X_test):
        """Predict response for ``X_test``. """
        # compute predictions ``predictions``
        return predictions

    def score(self, X_test, y_test):
        """Evaluate the estimator’s performance. """
        # compute performance measure ``score``
        return score
```

SKLEARN API: ESTIMATORS

Q: What estimators have we used?


scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
predicting a category
do you have labeled data
predicting a quantity
just looking
predicting structure
tough luck
classification
<100K samples
SGD Classifier
kernel approximation
Linear SVC
Text Data
Naive Bayes
KNeighbors Classifier
SVC
Ensemble Classifiers
clustering
number of categories known
<10K samples
KMeans
Spectral Clustering
GMM
MiniBatch KMeans
<10K samples
MeanShift
VBGMM
regression
<100K samples
SGD Regressor
few features should be important
Lasso
ElasticNet
RidgeRegression
SVR(kernel='linear')
SVR(kernel='rbf')
EnsembleRegressors
dimensionality reduction
Randomized PCA
<10K samples
Isomap
Spectral Embedding
LLE
kernel approximation
YES
NO
NOT WORKING
Back

EVALUATING CLASSIFIERS

WAYS OF CLASSIFYING

- Binary
- Multi-class
- Multi-label

PERFORMANCE MEASURES

How do we measure how correct a classification model is?

PERFORMANCE MEASURES

Example: Predict whether tumors are malignant or benign:

- ***Accuracy*: ** fraction of instances that are classified correctly
  - does not differentiate between malignant tumors that were classified as being benign, and benign tumors that were classified as being malignant.
- In some problems, the costs associated with all types of errors may be the same
- In this problem, failing to identify malignant tumors is likely more severe than failing to identify benign tumors as malignant

PERFORMANCE MEASURES

- **True positive**: correctly classifying a positive case
- **True negative**: correctly classifying a negative case
- **False positive**: a negative case incorrectly classified as positive
- **False negative**: a positive case incorrectly classified as negative

CONFUSION MATRIX

	Prediction		
Actual		1	0
	1	**TP**	**FP**
	0	**FN**	**TN**

PERFORMANCE MEASURES

- ***Accuracy*:** fraction of instances classified correctly

$$\mathrm{ACC} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})}$$

PERFORMANCE MEASURES

- ***Precision***: fraction of cases predicted to be positive that are actually positive

$$P = TP / (TP + FP)$$

PERFORMANCE MEASURES

- ***Recall***:
  *or True Positive Rate, (**TPR**)*
  *or **Sensitivity***

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

PERFORMANCE MEASURES

- ***Fall-out*:**
  or false positive rate ***(FPR)***
  or (1 - ***specificity****)*

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

ROC & AUC

ROC

The ***Receiver Operating Characteristic*** (ROC) is a graphical plot that illustrates the performance of a binary classifier system as its ***discrimination threshold*** is varied

ROC

The ***Receiver Operating Characteristic*** (ROC) is a graphical plot that illustrates the performance of a binary classifier system as its ***discrimination threshold*** is varied

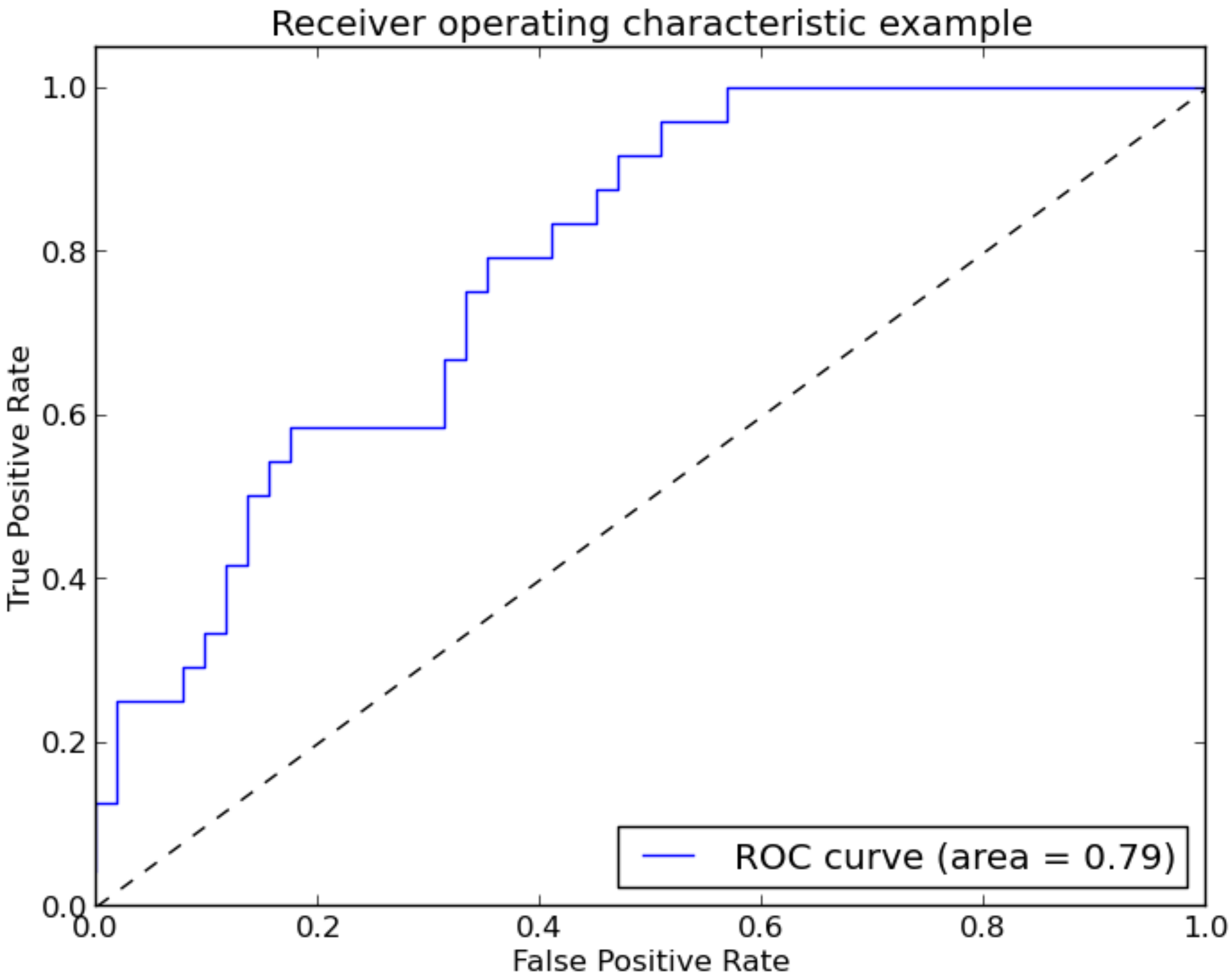
Create the curve by plotting the ***true positive rate*** against the ***false positive rate*** at various threshold settings

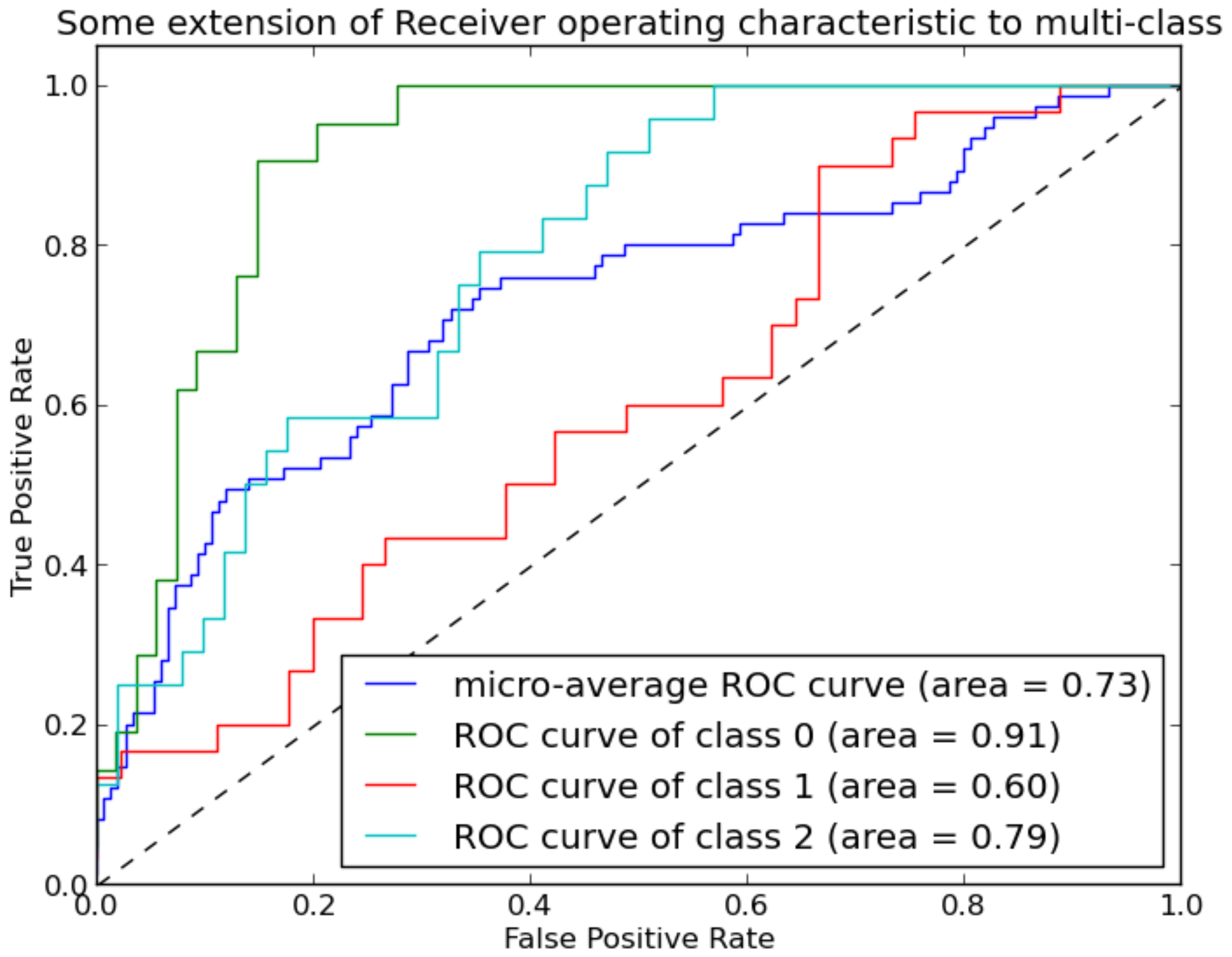
ROC: ORIGINS

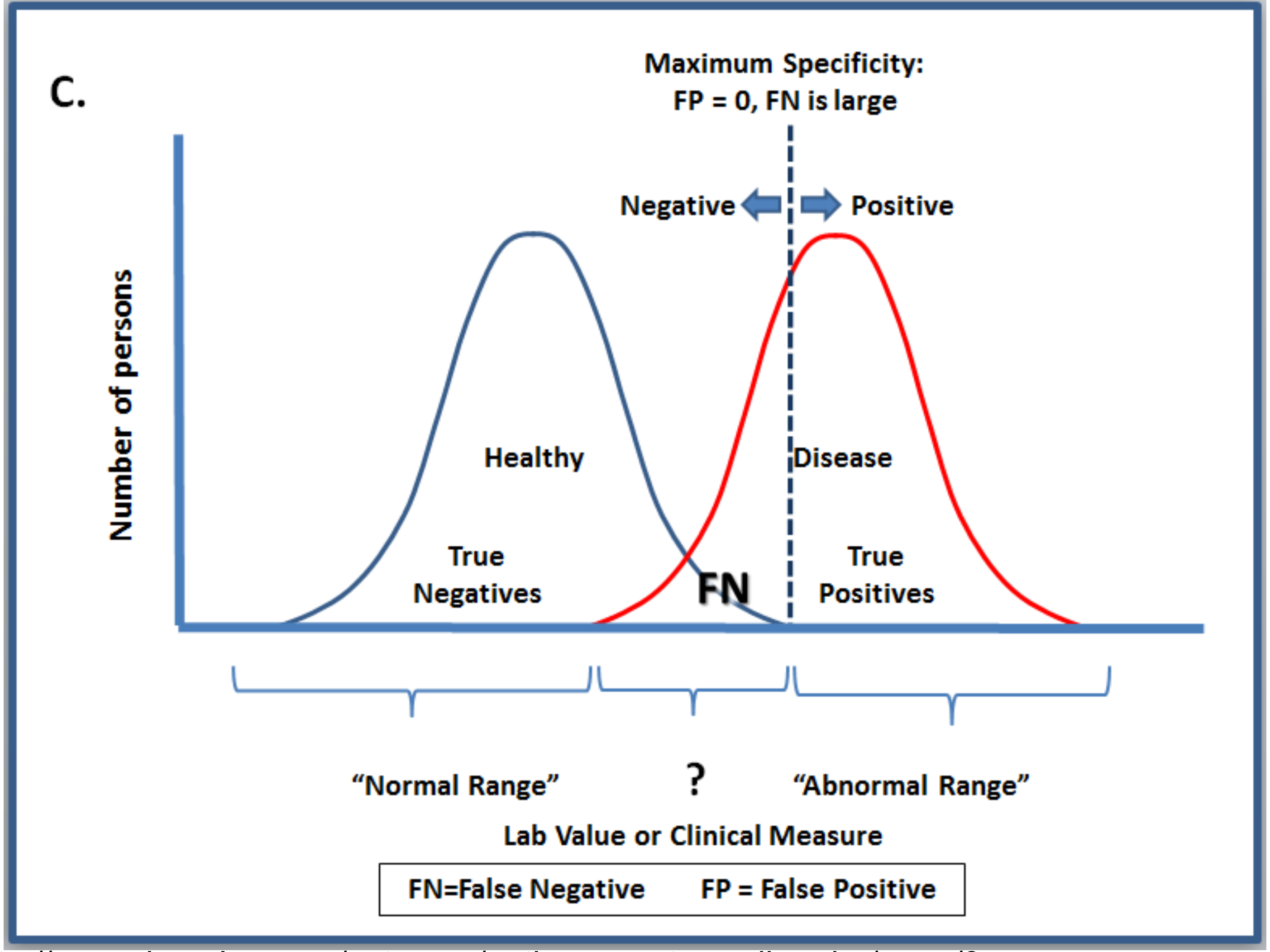
- Developed by “receiver operators” during WWII for radar-signal detection methodology (signal-to-noise), hence “Radar Receiver Operator Characteristic”)
- Used extensively in medical and psychological test evaluation
- More recently used in machine learning

ROC: USE-CASES

- Used to compare probabilistic forecasts of events or non-events
- Assess the tradeoff between ***sensitivity*** and ***specificity***
- Classify forecast probabilities into binary categories (0,1) based on probabilistic thresholds
- Compare detection ability of different experimental methods

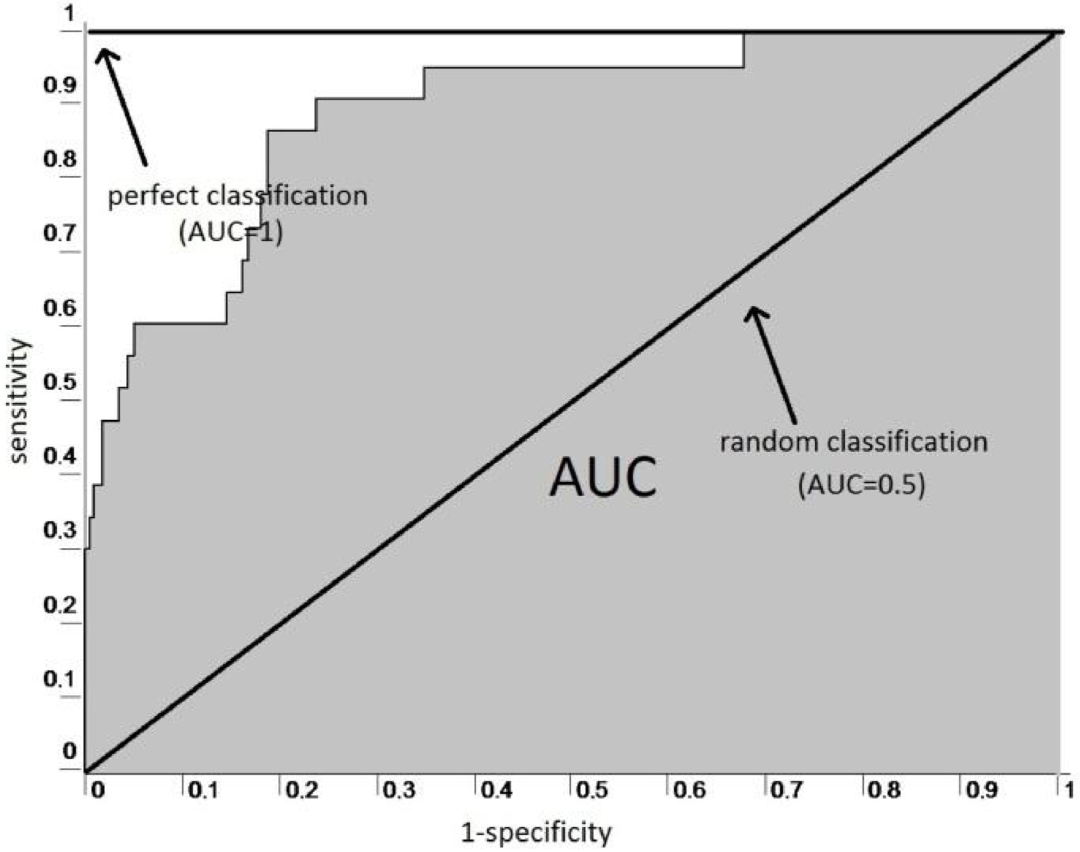

Receiver operating characteristic example
True Positive Rate
False Positive Rate
ROC curve (area = 0.79)
0.0
0.2
0.4
0.6
0.8
1.0


Some extension of Receiver operating characteristic to multi-class
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (area = 0.73)
ROC curve of class 0 (area = 0.91)
ROC curve of class 1 (area = 0.60)
ROC curve of class 2 (area = 0.79)



source: http://www.drcoplan.com/category/autism-spectrum-disorder/page/3

AUC: Area Under Curve



ROC Curve: What it Shows

- Tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).
- The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test
- The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
- The area under the curve (AUC) is a measure of test accuracy

Logistic Regression & Naive Bayes review

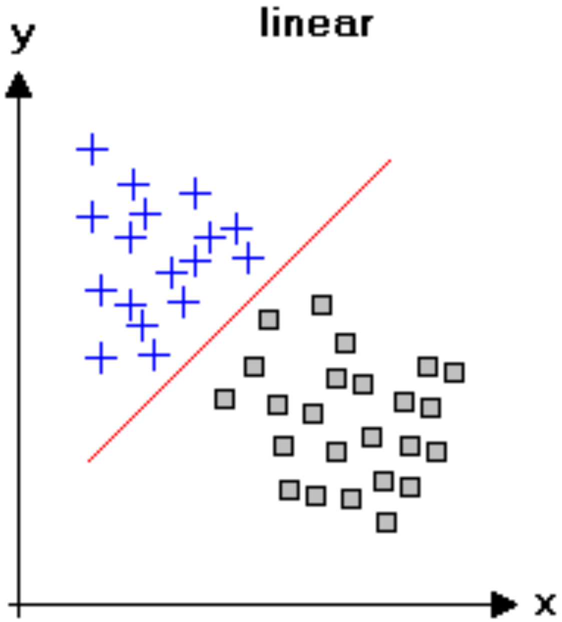
ADVANTAGES OF LOGISTIC REGRESSION

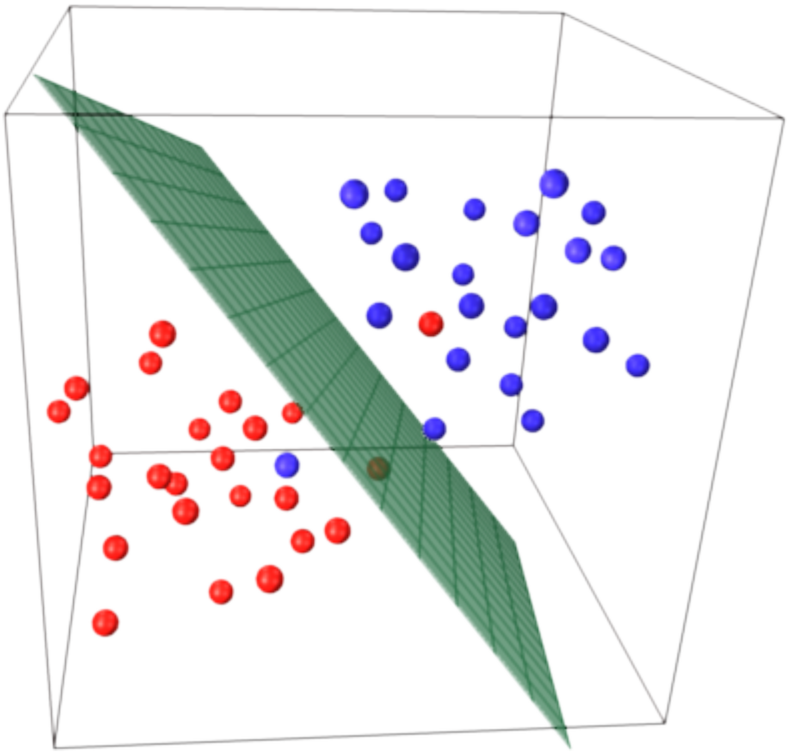
- Learns efficiently in high dimensional feature spaces
- Supports correlated explanatory variables
- Supports regularization
- Provides a confidence measure for predictions
- Supports ***online learning***

DISADVANTAGES OF LOGISTIC REGRESSION

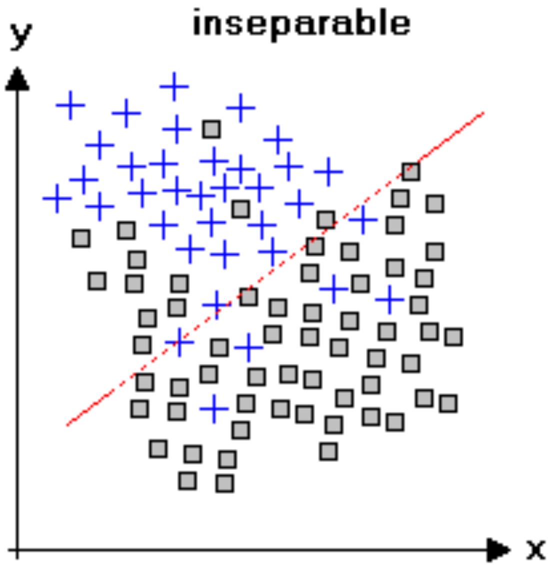
- Cannot effectively classify ***linearly inseparable*** data

LINEAR SEPARABILITY





LINEAR SEPARABILITY



ADVANTAGES OF Naive Bayes

- Easy to implement
- Simple, surprisingly effective model in many datasets
- Requires small amount of training data to estimate parameters
- Converges more quickly than Logistic Regression
- Online learning capable

DISADVANTAGES OF Naive Bayes

- Because of conditional independence assumption, cannot learn interactions between features

YOU MIGHT ALSO CONSIDER

- LINEAR SUPPORT VECTOR MACHINES
- NAIVE BAYES
- PERCEPTRONS
- LINEAR/QUADRATIC DISCRIMINANT ANALYSIS

IF THOSE DON'T WORK, YOU MIGHT TRY

- NON-LINEAR SUPPORT VECTOR MACHINES
- RANDOM FORESTS
- K-NEAREST NEIGHBORS (κNN)
- MULTI-LAYER PERCEPTRONS

INTRO TO DATA SCIENCE