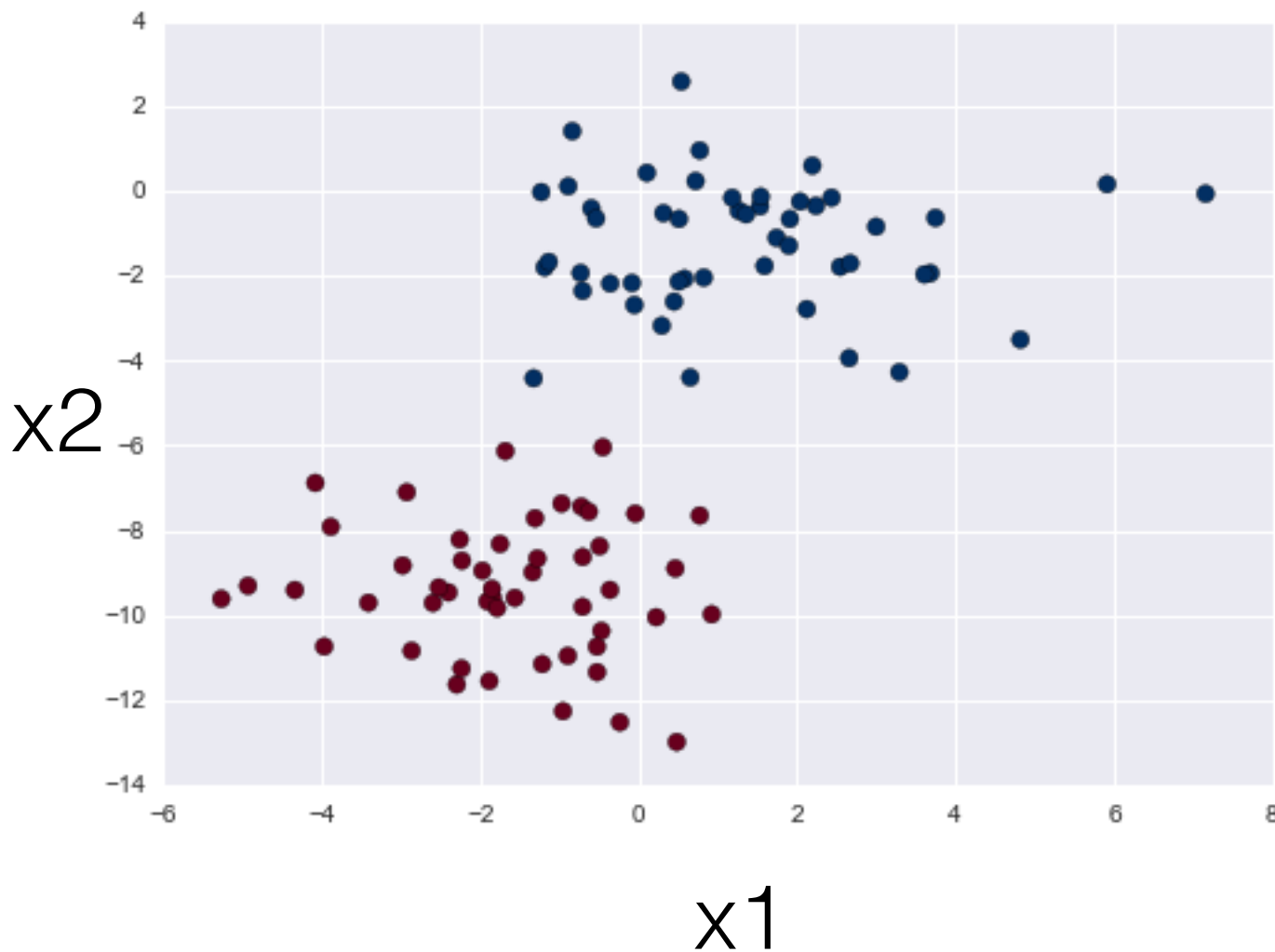# CLASSIFICATION II

11.23.2020

CLASSIFIERS

x2

x1

* a classifier is a function that guesses the "class" of a datapoint based on its features

* here "class" is **red** or **blue** and each point has 2 features (x1 and x2)

CLASSIFIERS *READING*

* **Chapter 17** of *Inferential Thinking*: https://www.inferentialthinking.com/chapters/17/Classification.html

* **Chapter 5** of *PDSH*: https://jakevdp.github.io/PythonDataScienceHandbook/05.00-machine-learning.html (in particular 5.07 and 5.05)

CLASSIFIERS

* classification shares *a lot* in common with regression

* (one of the classification methods we'll be looking at today is called *Logistic Regression*)

CLASSIFIERS

* Today we'll test two different classifiers from scikit-learn:

  * **Logistic Regression** (linear)

  * **Support Vector Classifier** (nonlinear)

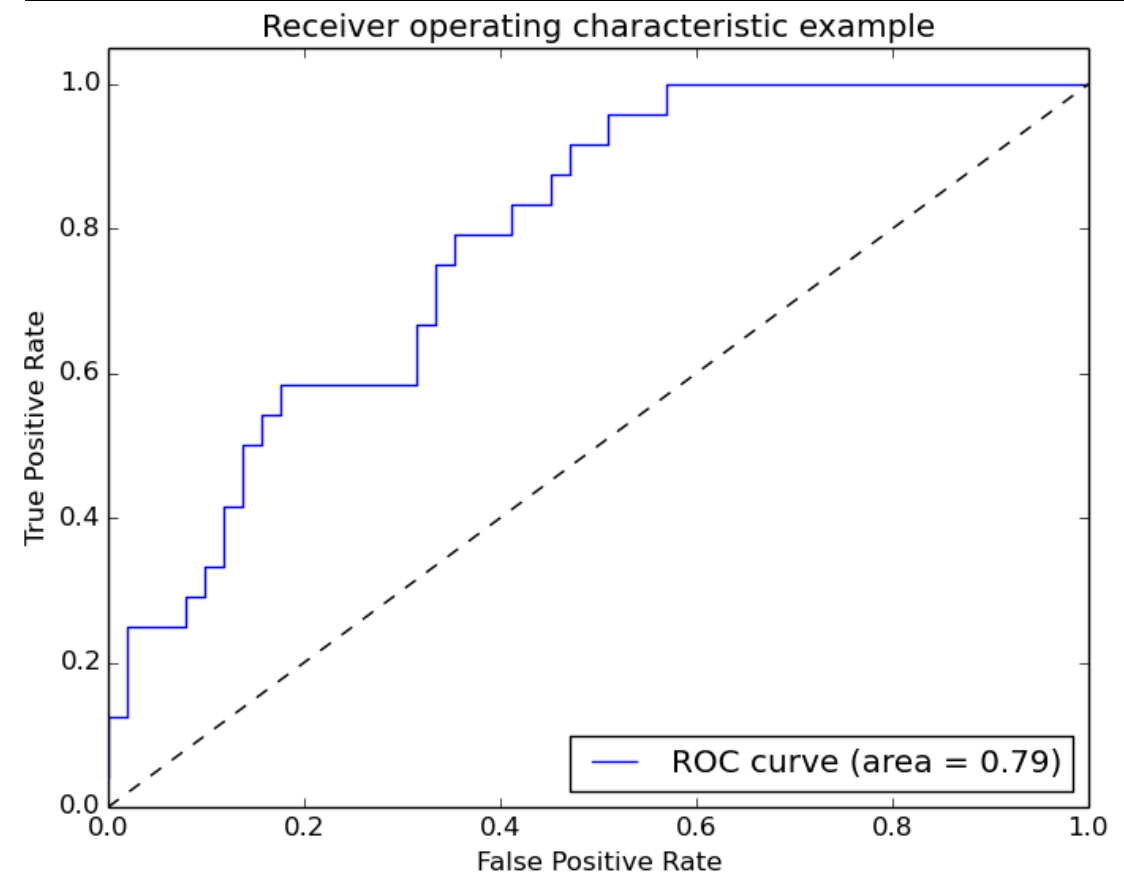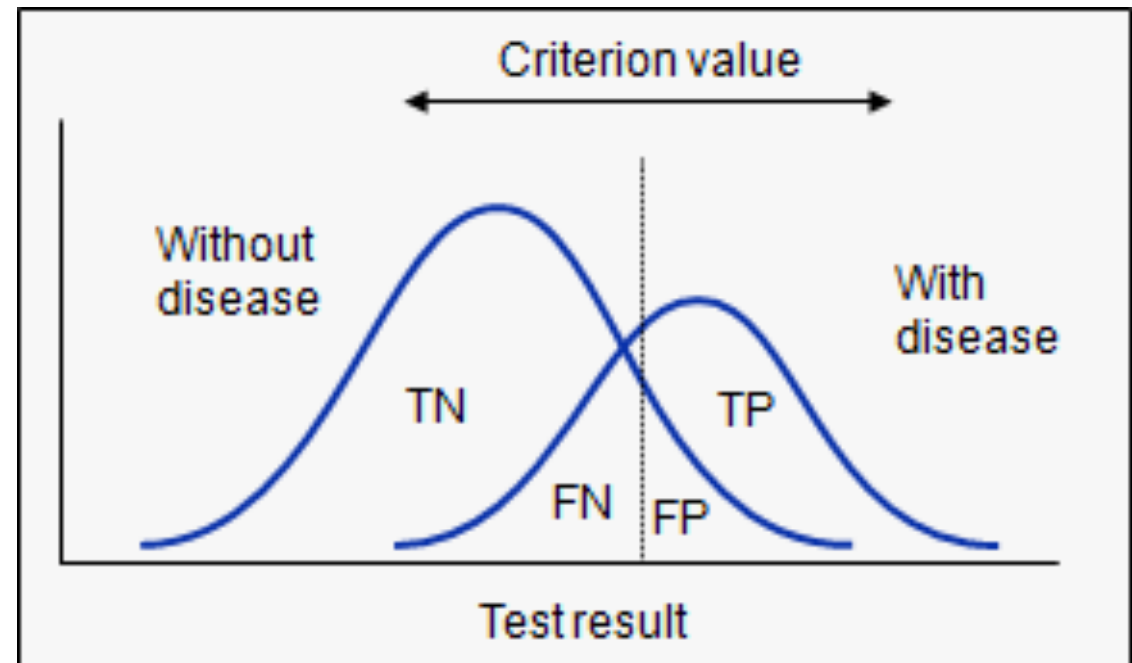* (see also https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html)

ROC ANALYSIS

* to measure classifier effectiveness, we can use **ROC analysis**

* we measure the true positive rate (tpr) and false positive rate (fpr) for different "criterion values", then plot
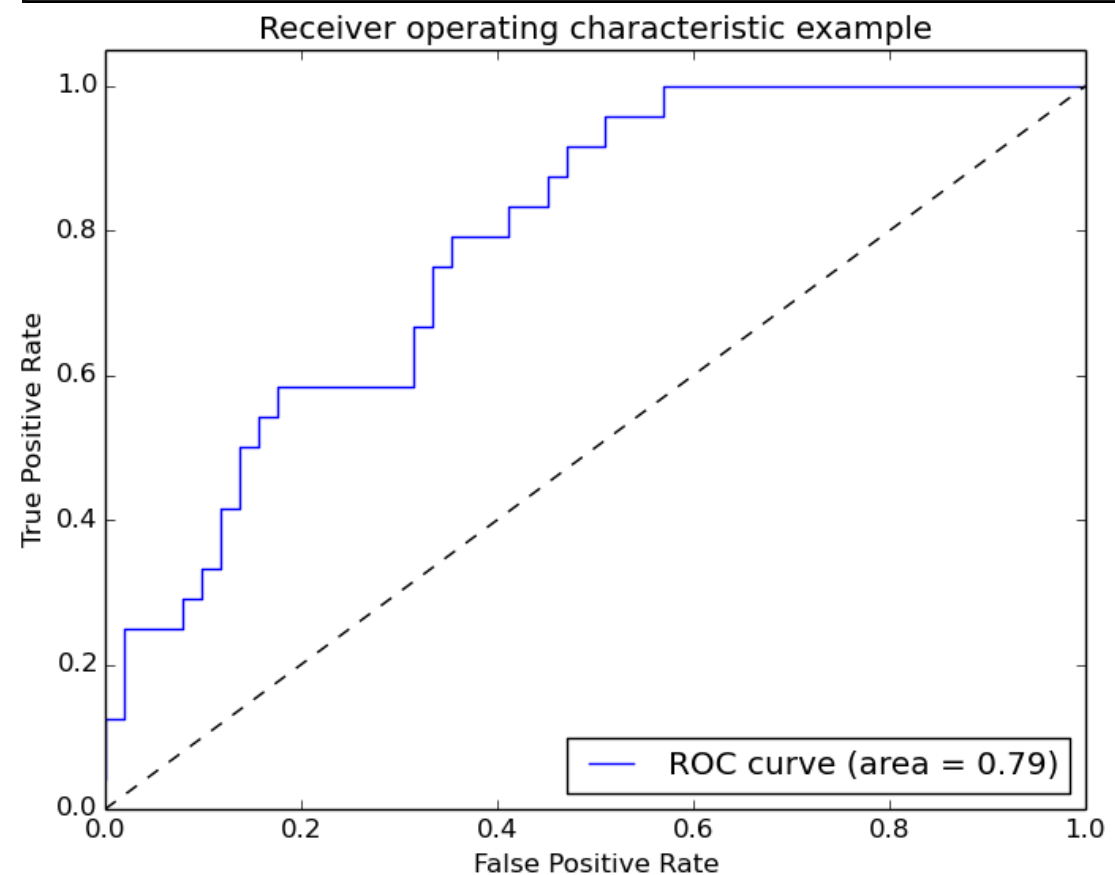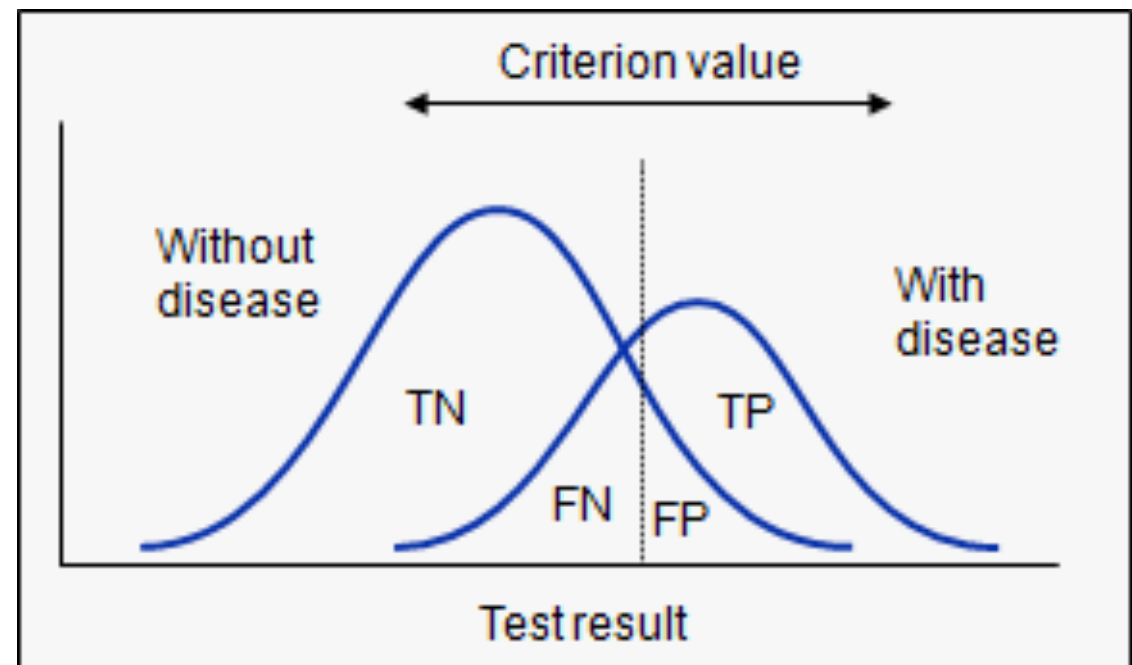
ROC ANALYSIS

* you then plot the TP vs. FP rate to get an ROC curve

NEXT TIME

* on *NEXT* Monday we'll begin talking about **CLUSTERING** and **DIMENSIONALITY REDUCTION**

THE END