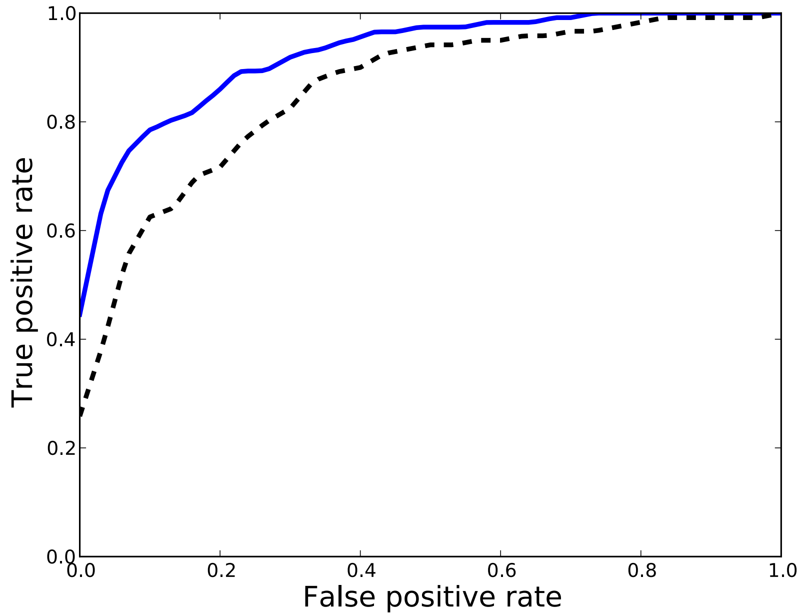
Measuring classifier accuracy



Measures of classifier performance

Classifier performance can be summarized by a table known as the confusion matrix or contingency table:

true labels

predicted labels:

	-1	1
-1	1439	61
1	62	1438

Measures of classifier performance

Let's take a close look at the confusion matrix:

predicted labels:

true labels		-1	1
	-1	1439	61
	1	62	1438

How do we compute error from the confusion matrix?

Measures of classifier performance

For binary classification problems it is customary to express the contingency table as:

```
true labels    predicted labels:

                      -1    1
               -1     TN    FP
                1     FN    TP
```

TP – number of true positives

TN – number of true negatives

FP – number of false positives

FN – number of false negatives

Measures of classifier performance

For binary classification problems it is customary to express the contingency table as:

true labels	predicted labels:		
	-1	1	
-1	TN	FP	Neg = TN+FP
1	FN	TP	Pos = TP+FN

True positive rate/sensitivity/recall: TP / Pos

True negative rate/specificity: TN / Neg

False positive rate: FP / Neg

Precision: TP / (TP + FP)

Measures of classifier performance

Suppose you have a dataset with very few positive examples compared to negative examples (unbalanced data)

A classifier that classifies every example as negative would still attain high accuracy (this is called the majority class classifier).

Need alternative measures of accuracy!

The choice of classification threshold

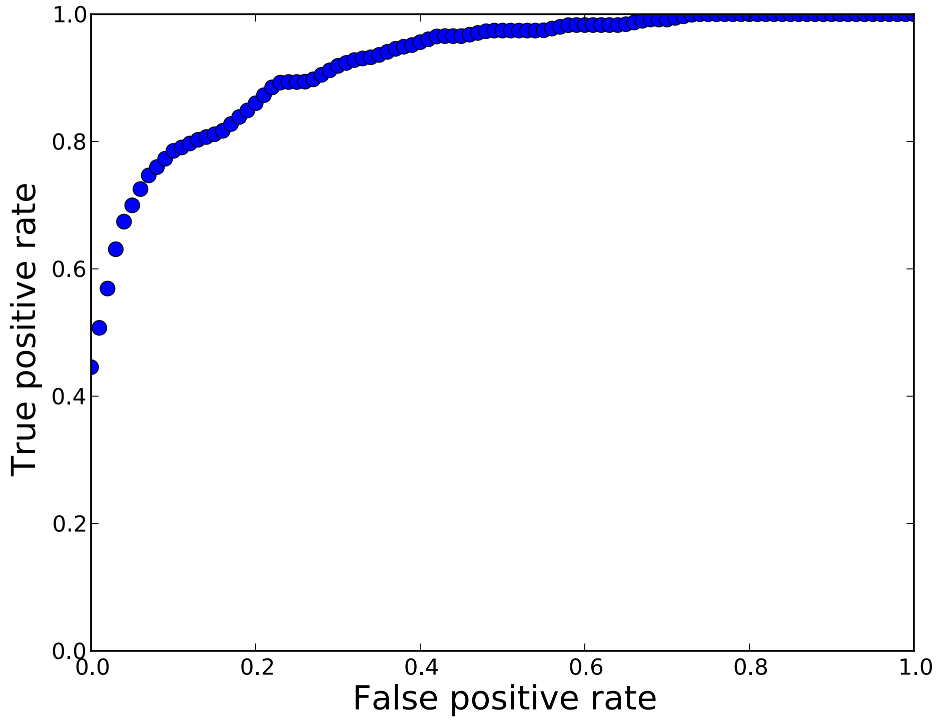
All the classifiers we will study provide a scoring function whose magnitude indicates how sure we are it belongs to a given class.
For example: $\mathbf{w}^{\mathsf{T}}\mathbf{x} + b$

The choice of the threshold is arbitrary, and in a given application we may prefer to ignore positive predictions that have small scores

To have a view of classifier performance that is independent of the choice of threshold we consider the ROC curve.

ROC curves

The ROC curve is a plot of the true positive rate as a function of false positive rate as you vary the classification threshold



How does the ROC curve of a perfect classifier look like?

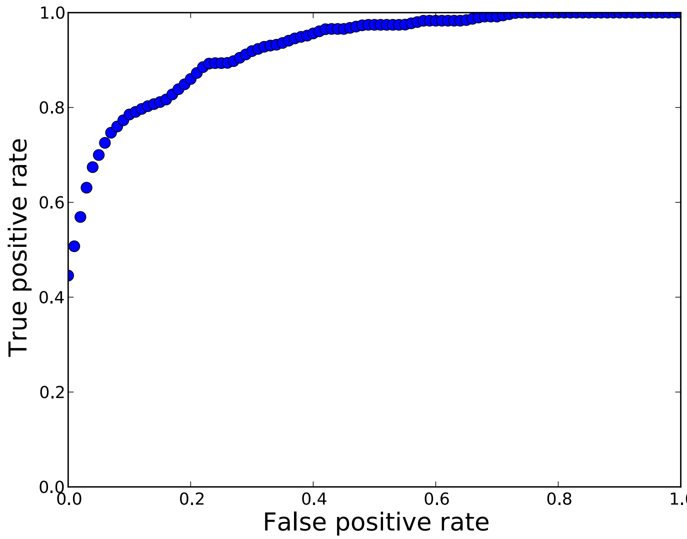
For a random classifier?

ROC curve computed on the heart disease dataset from the UCI repository

ROC curves

An ROC curve is often summarized by the area under the curve (AUC).

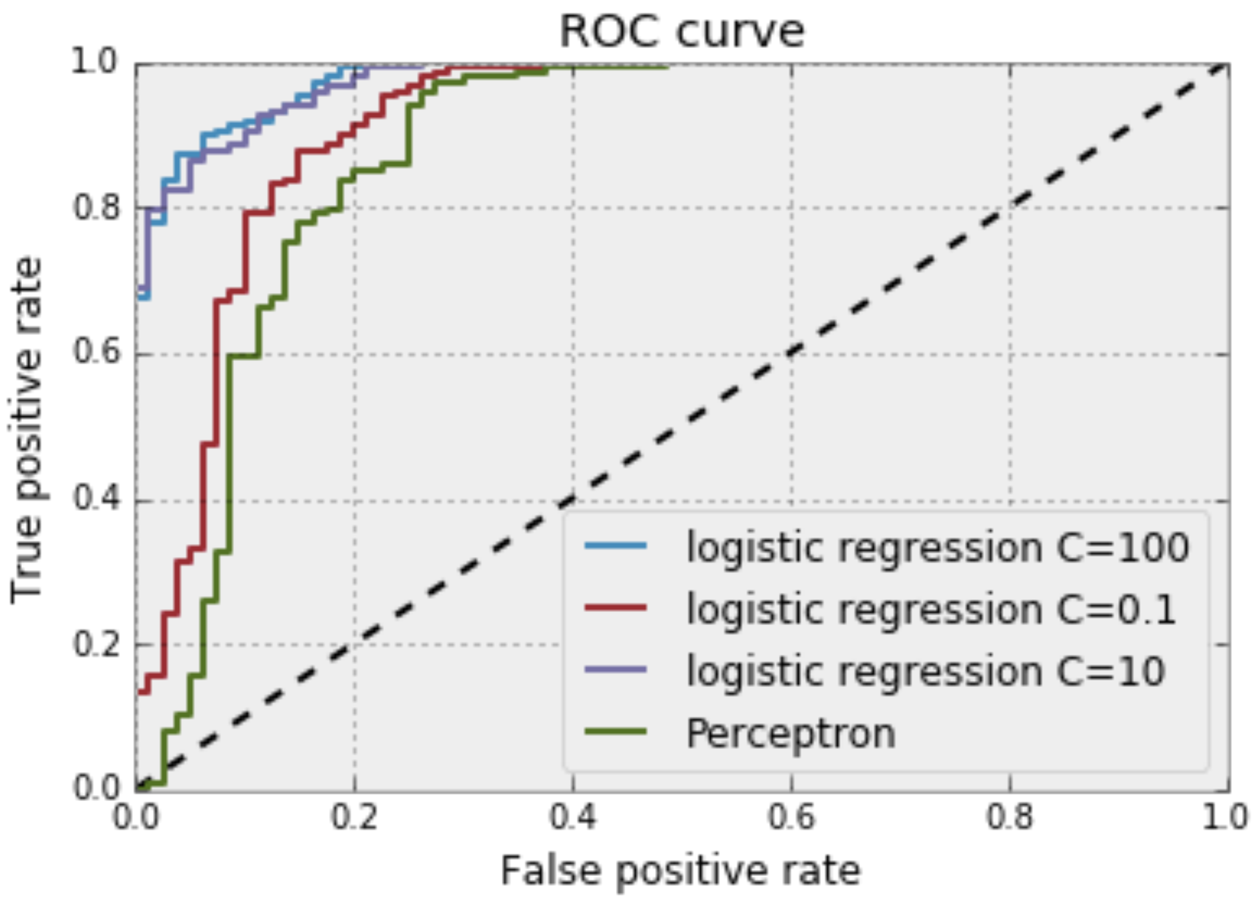
AUC = 0.92



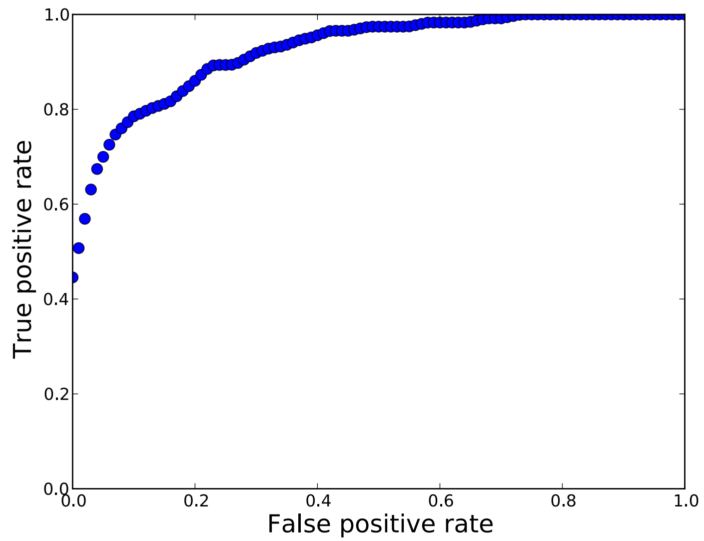
AUC is essentially the probability that a positive example will get a higher score than a negative example

ROC curves

ROC curves are a nice way of comparing classifiers:



ROC curves



Additional comments:

- Applicable only to binary classification.
- REC curves: similar idea for regression problems (see assignment 3)
- ROC curves can be misleading when data is highly imbalanced, in which case it is possible to focus on the beginning of the curve

Precision Recall Curves

Precision-Recall curves are another good alternative: