EVALUATING CLASSIFIERS

LECTURE 5
SECTION 1
JUNE 9th

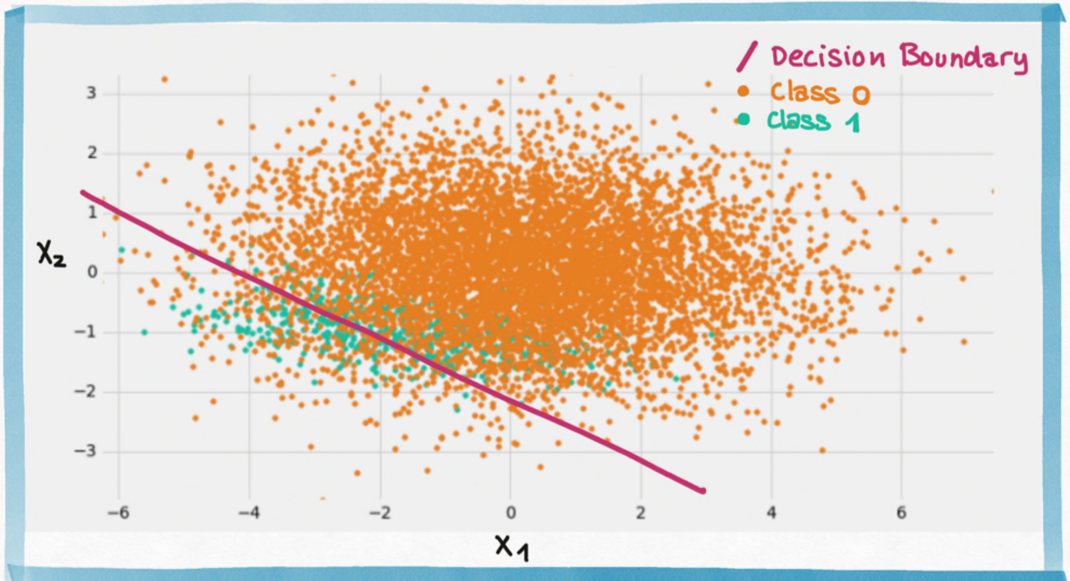
IACS
INSTITUTE FOR APPLIED
COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY

UNIVERSITY of
RWANDA

EVALUATING CLASSIFIERS UNDER CLASS IMBALANCE

CLASS IMBALANCE:

We've seen that when the classes in our training data are extremely unbalanced we need to look at the confusion matrix rather than the accuracy:



Train Accuracy: 95%
Test Accuracy: 94%

Confusion Matrix:

	$\hat{y}=1$	$\hat{y}=0$
$y=1$	40	60
$y=0$	40	1960

Our performance on the minor class is poor. What should we do?

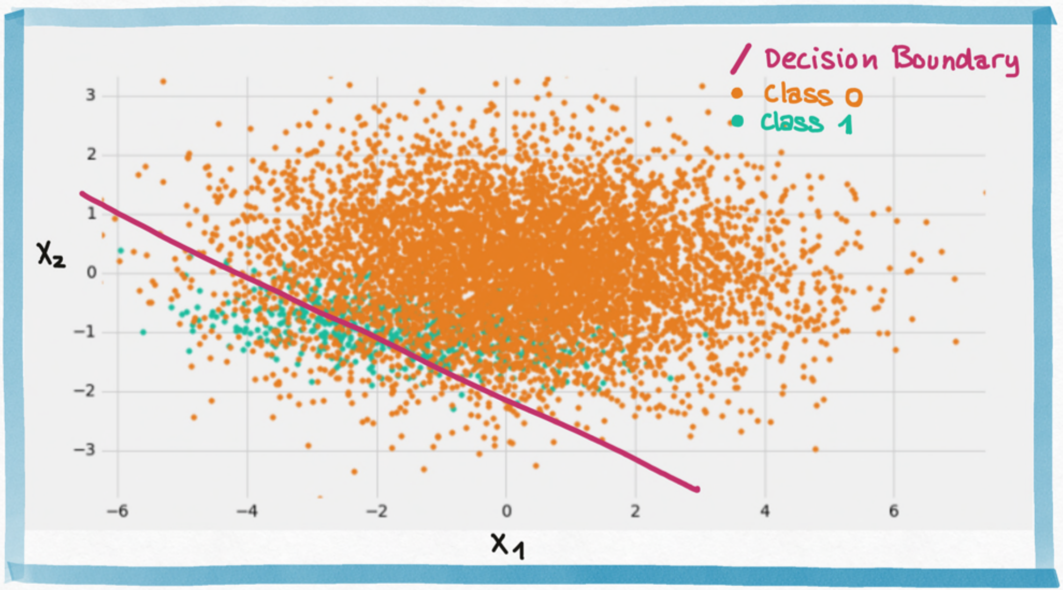
The answer depends on our application and on which class we care more about.

POSITIVE AND NEGATIVE CLASSIFICATION RATES

FALSE AND TRUE POSITIVE RATES:

If our classifier is being used to predict whether or not an accused person is guilty of a crime, perhaps its performance is acceptable, since very few innocent people are wrongly accused.

This is the false positive rate: $P(\hat{y}=1 \mid y=0) = \dfrac{\#\text{incorrect } \hat{y}=1 \text{ pred's}}{\# \, y=0} = \dfrac{40}{2000}$



Train Accuracy: 95%
Test Accuracy: 94%

Confusion Matrix:

	$\hat{y}=1$	$\hat{y}=0$
$y=1$	40	60
$y=0$	40	1960

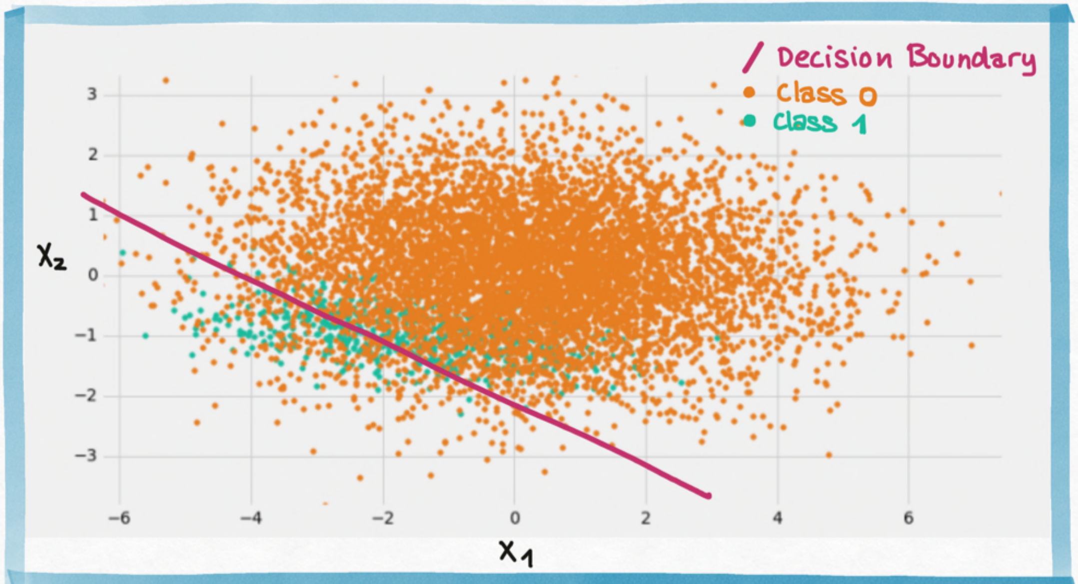
The true positive rate is: $P(\hat{y}=1 \mid y=1) = \dfrac{\#\text{ correct } \hat{y}=1 \text{ pred's}}{\# \, y=1} = \dfrac{40}{100}$.

In this case, the false positive rate is more important.

FALSE AND TRUE NEGATIVE RATES:

If our classifier is being used to diagnose patients with a fatal disease, then its performance would not be acceptable, since it falsely classifies most sick patients as disease free.

This classifier has a high false negative rate: $p(\hat{y}=0 \mid y=1) = \frac{60}{100}$.



Train Accuracy: 95%

Test Accuracy: 94%

Confusion Matrix:

	$\hat{y}=1$	$\hat{y}=0$
$y=1$	40	60
$y=0$	40	1960

The true negative rate is: $p(\hat{y}=0 \mid y=0) = \frac{1960}{2000}$, which is high, but in this case we care more about the false negative rate.

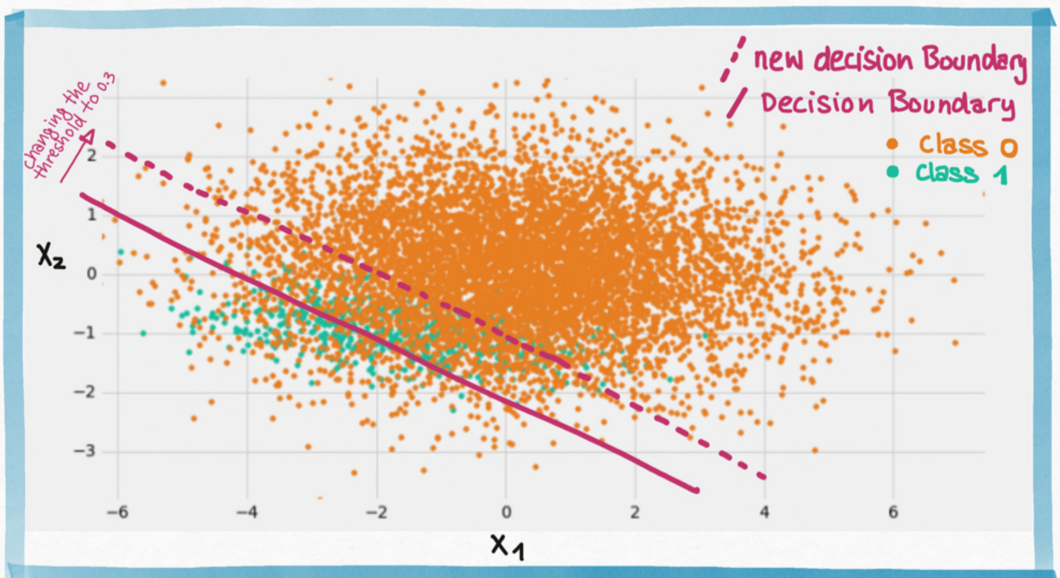
CHANGING THE CLASSIFICATION THRESHOLD

ADJUSTING THE DECISION THRESHOLD:

If we want to change the false positive or negative rates, we need to change the threshold for labeling a point $\hat{y}=1$.

To lower the false negative rate, we can decide to label $\hat{y}=1$ if $\sigma(f_w(x))=0.3$ instead of 0.5, for example. This lowers the threshold for predicting 1.

To lower the false positive rate, we can decide to label $\hat{y}=1$ if $\sigma(f_w(x))=0.7$ instead of 0.5, for example. This raises the threshold for predicting 1.



Train Accuracy: 74%

Test Accuracy: 73%

Confusion Matrix:

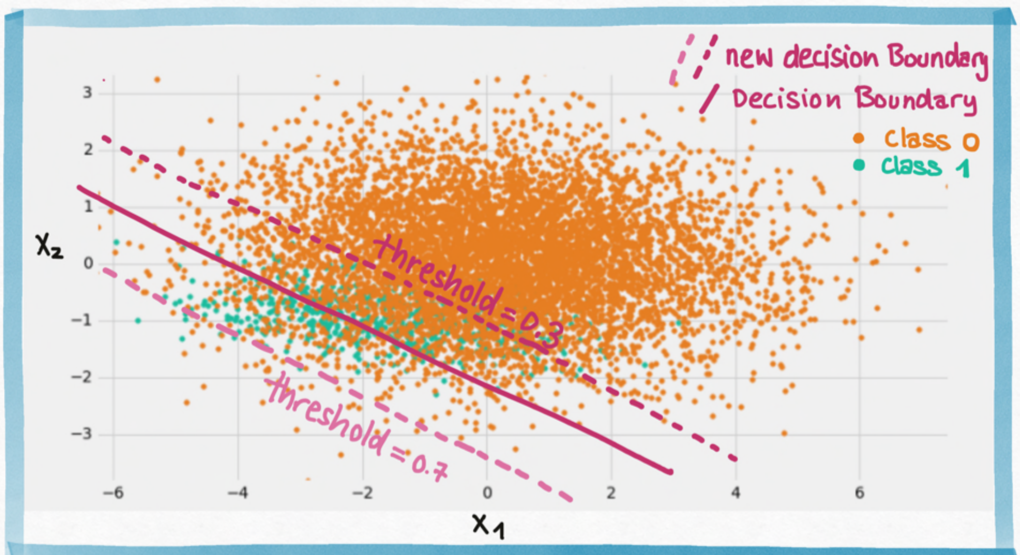
	$\hat{y}=1$	$\hat{y}=0$
$y=1$	80	20
$y=0$	600	1400

TRADE-OFFS BETWEEN RATES:

Lowering the classification threshold lowers the false negative rate, and raises both the true positive and false positive rates.

Raising the classification threshold lowers the false positive rate, and raises both the true negative and false negative rates.

How do we manage this trade-off?



NEW EVALUATION METRICS

THE RECIEVING OPERATING CHARACTERISTIC:

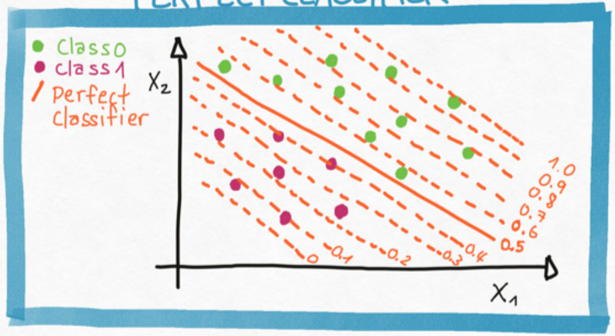
To understand the trade-offs when we change the classification threshold, we plot the false positive rate (FPR) and the true positive rate (TPR) for a range of thresholds, $t = 0, 0.1, \ldots, 1$. This is called the Receiving Operating Characteristic (ROC) curve.

For a perfect classifier, when t is between 0 and 0.5, the true positive rate is 1 (i.e. the false negative rate is zero), the false positive rate decrease as t increases. When t is between 0.5 and 1, the false positive rate is zero, and the true positive rate decreases as t increases.

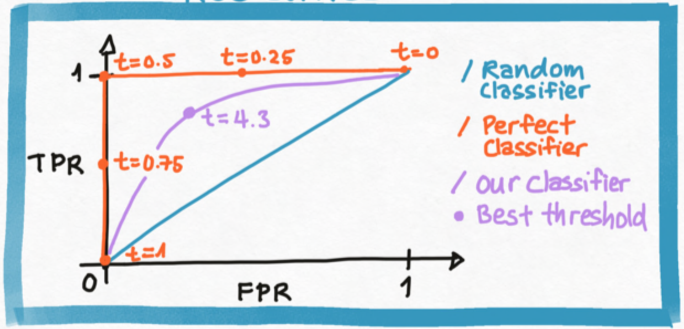
For a random classifier TPR is alway equal to FPR, regardless of t.

These two comparisons help us understand any classifer: we want to be like the perfect classifier and unlike the random classifier.

PERFECT CLASSIFIER



ROC CURVES



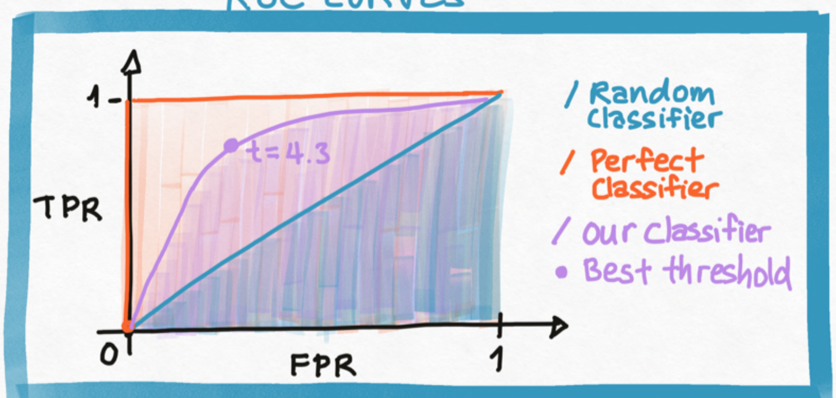
AREA UNDER THE ROC CURVE:

The ROC curve allows us to find the classification threshold that gives the best FPR and TPR trade-off (ie where FPR is lowest and TPR is highest.

But if all we are interested in is comparing our classifier against a perfect classifier and a random classifier, then we want to summarize the information in the ROC curve.

We summarize the ROC by computing the Area Under the ROC Curve (AUC). For a perfect classifier the AUC is 1, the random classifier has AUC of 0.5. We want our classifier to have AUC as close to 1 as possible.

ROC CURVES



AUC'S