Evaluating Models with ROC Curves

- Receiving Operating Characteristic, or ROC, is a visual way for inspecting the performance of a binary classification algorithm.
- In particular, it's comparing the rate at which your classifier is making correct predictions (*True Positives* or TP) and the rate at which your classifier is making false alarms (*False Positives* or FP).
- When talking about True Positive Rate (TPR) or False Positive Rate (FPR) we're referring to the definitions below:

$$\text{TPR} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$$

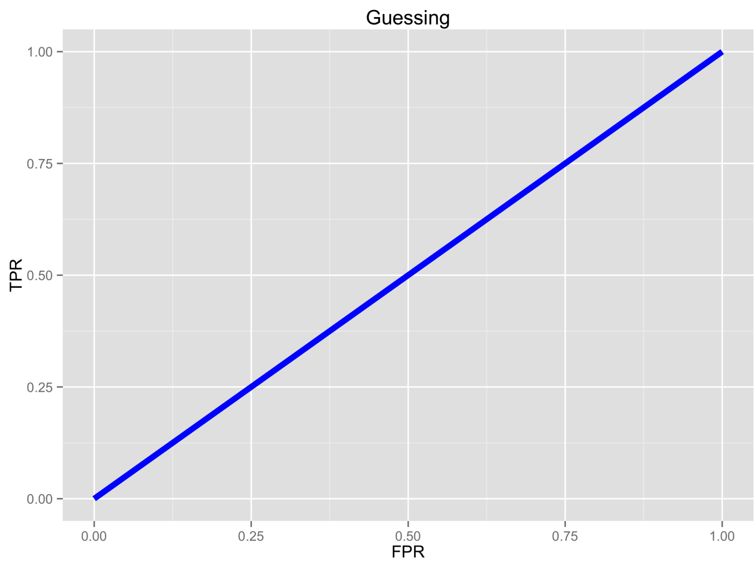
Remark *True Positives Rates and True Negatives Rates referred to as Sensitivity and Specificity.*

Background

- ROC curves were first used during WWII to analyze radar effectiveness.
- In the early days of radar, it was sometimes hard to tell a large bird from an incoming airplane.
- The British Ministry of Defence pioneered using ROC curves to optimize the way that they could rely to radar for detect approaching Luftwaffe airplanes.

Scenarios: Guessing at Random

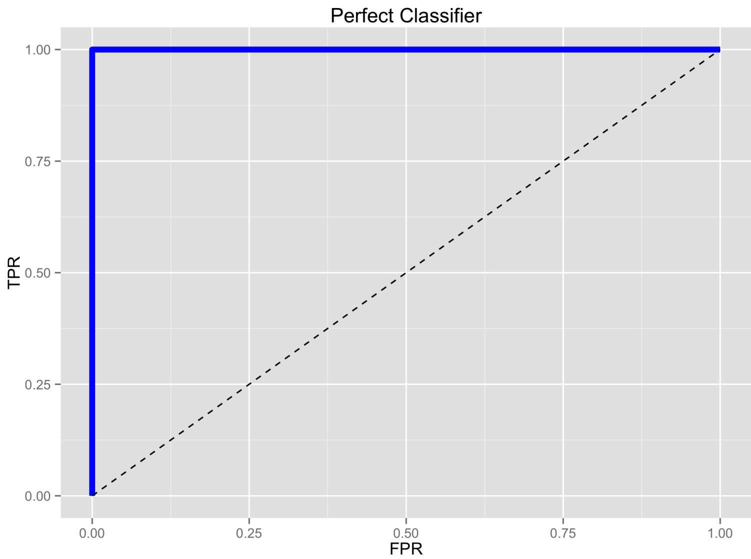
- The first example is the simplest: a diagonal line.
- A diagonal line indicates that the classifier is just making completely random guesses.
- Since your classifier is only going to be correct 50% of the time, it stands to reason that your TPR and FPR will also be equal.



Often, ROC charts will include the random ROC curve to provide the user with a benchmark for what a naive classifier would do.
Any curves above the line are better than guessing, while those below the line, you would be better off guessing.
For review: The Area Under the Curve (AUC) is 0.500.

A Perfect Classifier

- A perfect classifier will yield a perfect trade-off between TPR and FPR (meaning you'll have a TPR of 1 and an FPR of 0).
- In that case, your ROC curve looks something like this.

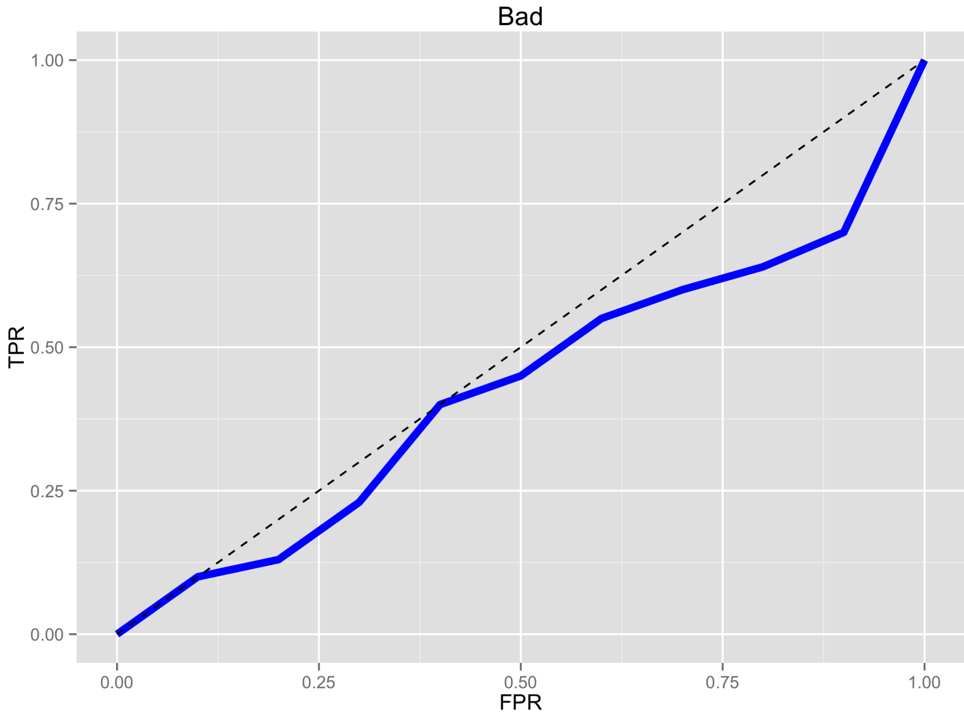


Important: The better your classifier, the more closer the curve will be to the top left corner.

For review: Note the "random curve" is included as a benchmark as a dotted line.
The Area Under the Curve (AUC) is 1.

Worse than guessing

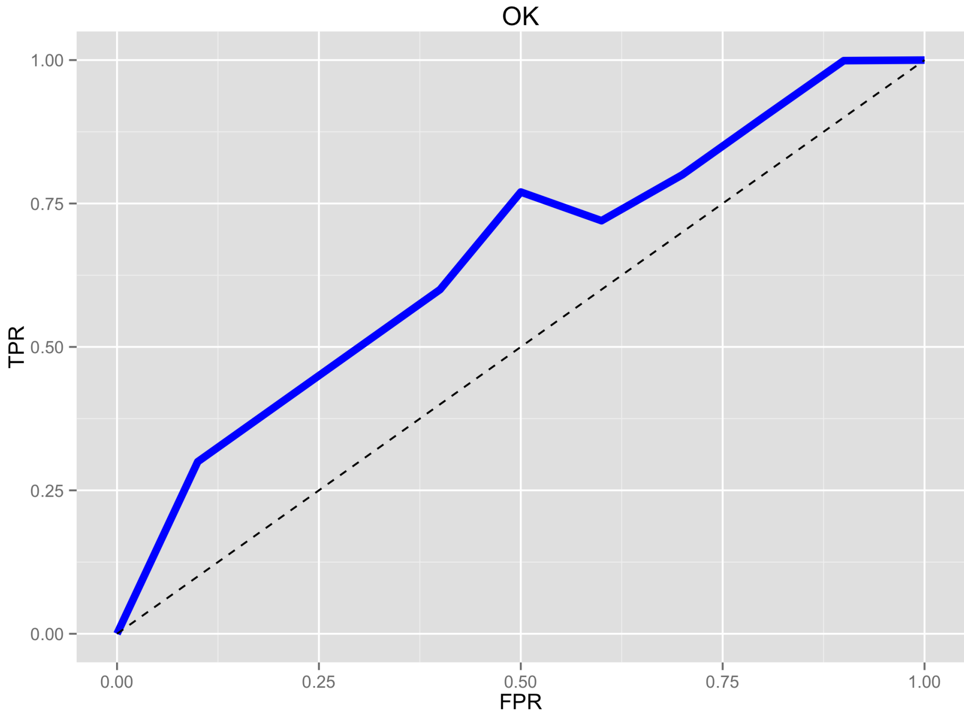
A bad classifier (i.e. something that's worse than guessing) will appear below the random line.



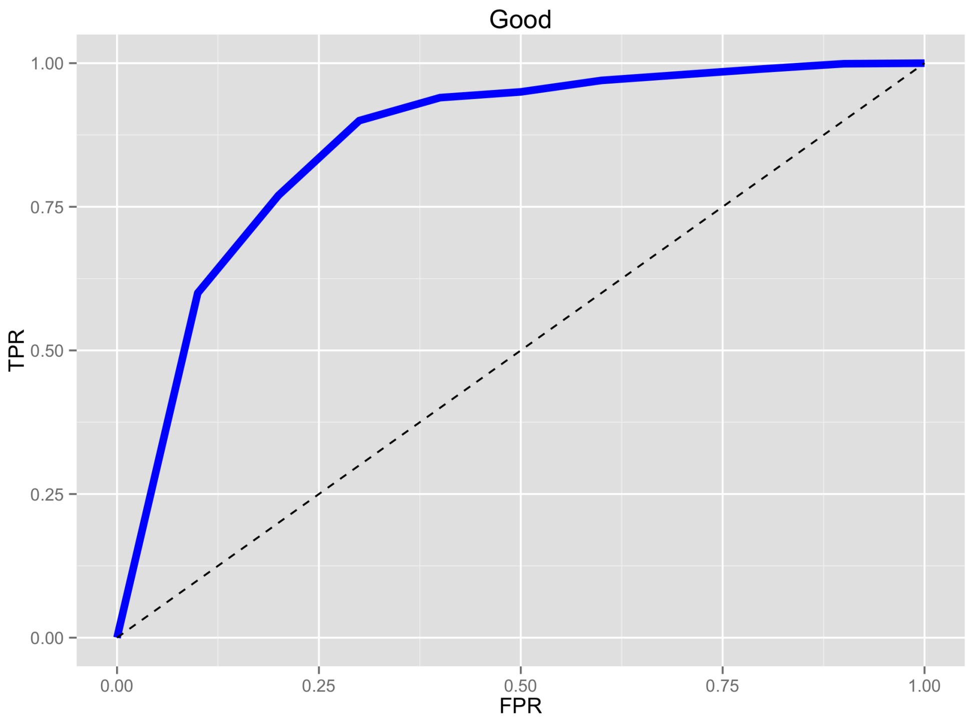
There have been several instances of a "prediction system" underperforming "guessing at random".

Better than guessing

A much more interesting activity is attempting to decipher the difference between an "OK" and a "Good" classifier. The chart below shows an example of a very mediocre classifier.



Pretty good



Area under the curve (AUC)

There is an aggregate metric to determine how good the prediction system is: AUC or Area Under the Curve.
The AUC is the amount of space underneath the ROC curve

- $AUC = 0$: Perfectly Bad
- $AUC < 0.5$: Worse than guessing at random
- $AUC = 0.5$: same as guessing at random
- $AUC > 0.5$: Good. better than guessing at random
- $AUC = 1$: Perfectly Good

Comparing AUC values is useful when comparing different models, as we can select the model with the high AUC value, rather than just look at the curves.