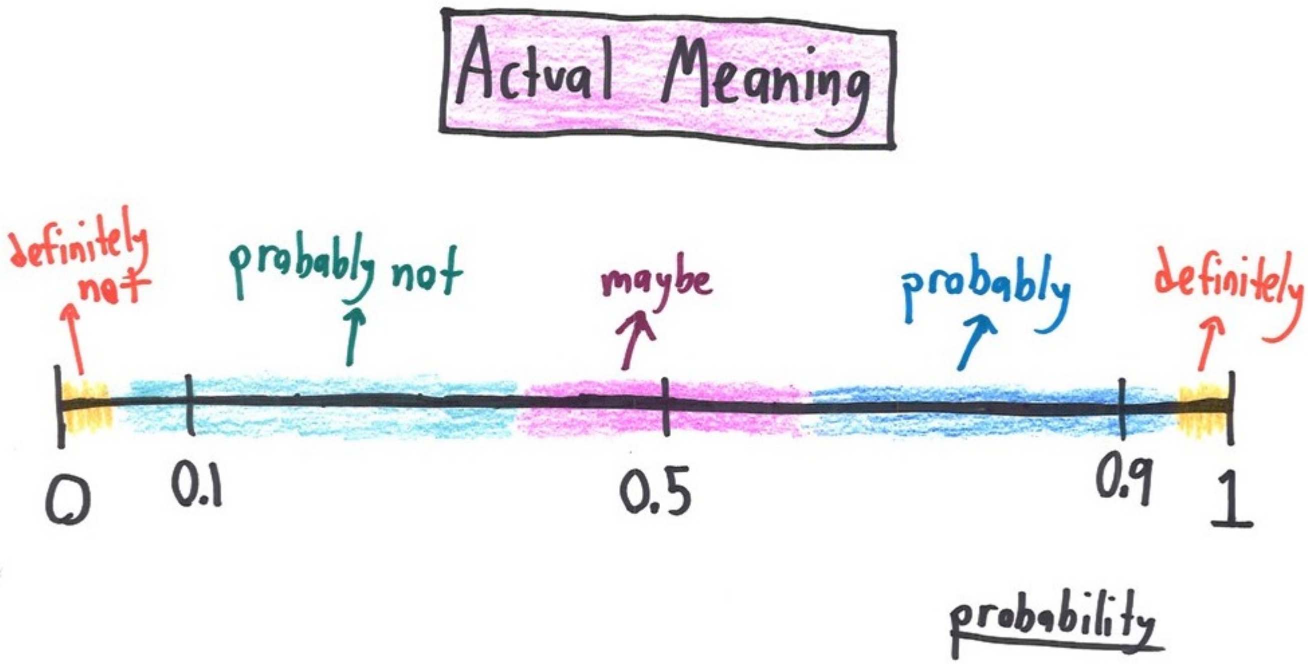
Loss Functions

Lyle Ungar

Loss functions come from decision making

- **Don't confuse *probability* with loss**
  - One can optimize a model for probability and then use the probability in a decision rule
  - Or just directly optimize the loss resulting from a decision rule

What does probability mean?

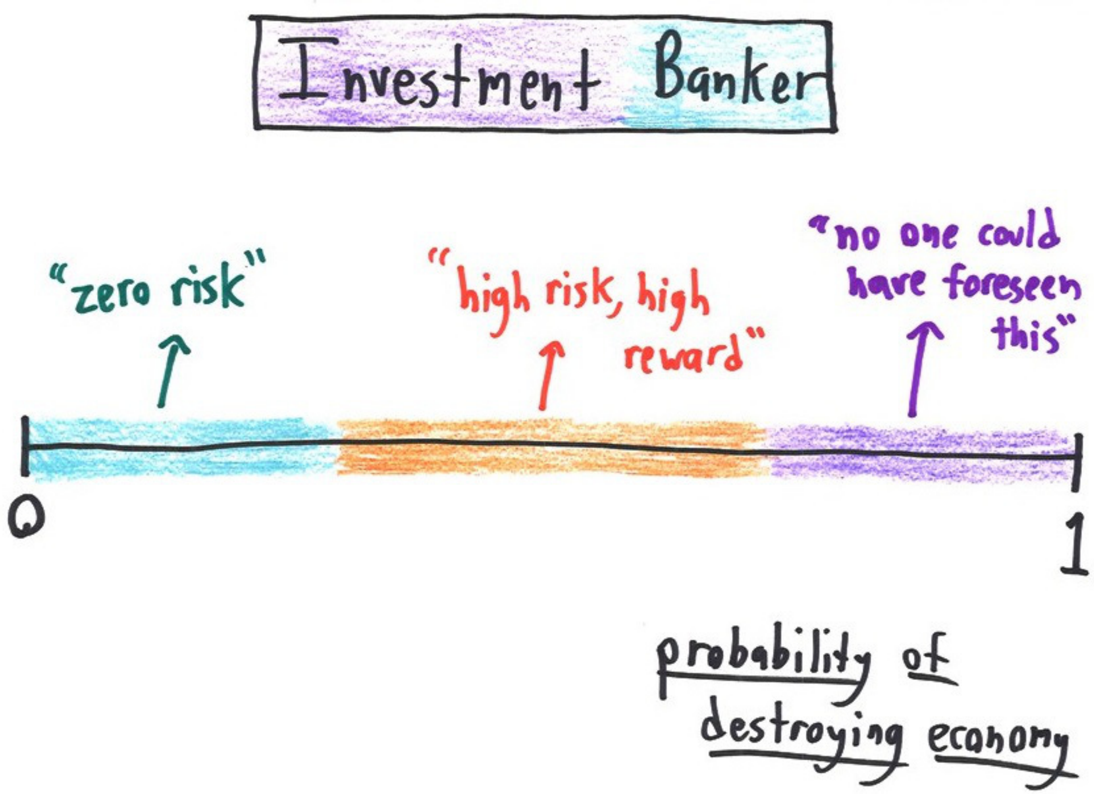


http://mathwithbaddrawings.com/2015/09/23/what-does-probability-mean-in-your-profession/

What does probability mean?

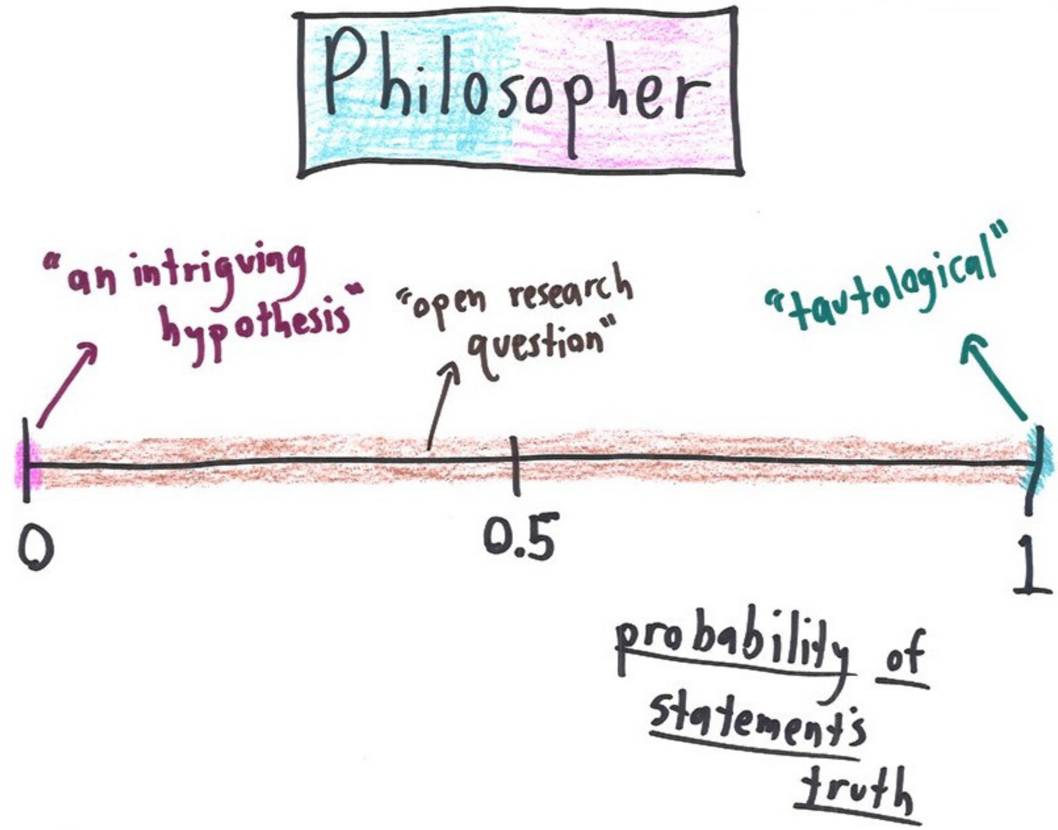
http://mathwithbaddrawings.com/2015/09/23/what-does-probability-mean-in-your-profession/

What does probability mean?



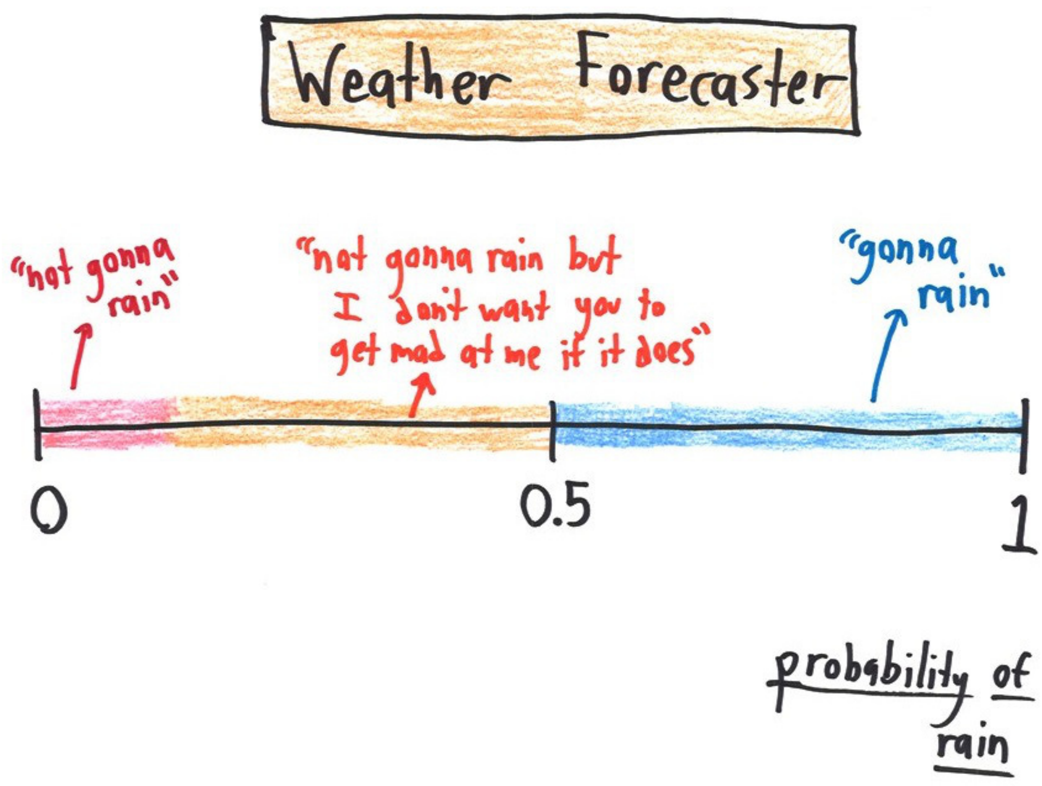
http://mathwithbaddrawings.com/2015/09/23/what-does-probability-mean-in-your-profession/

What does probability mean?



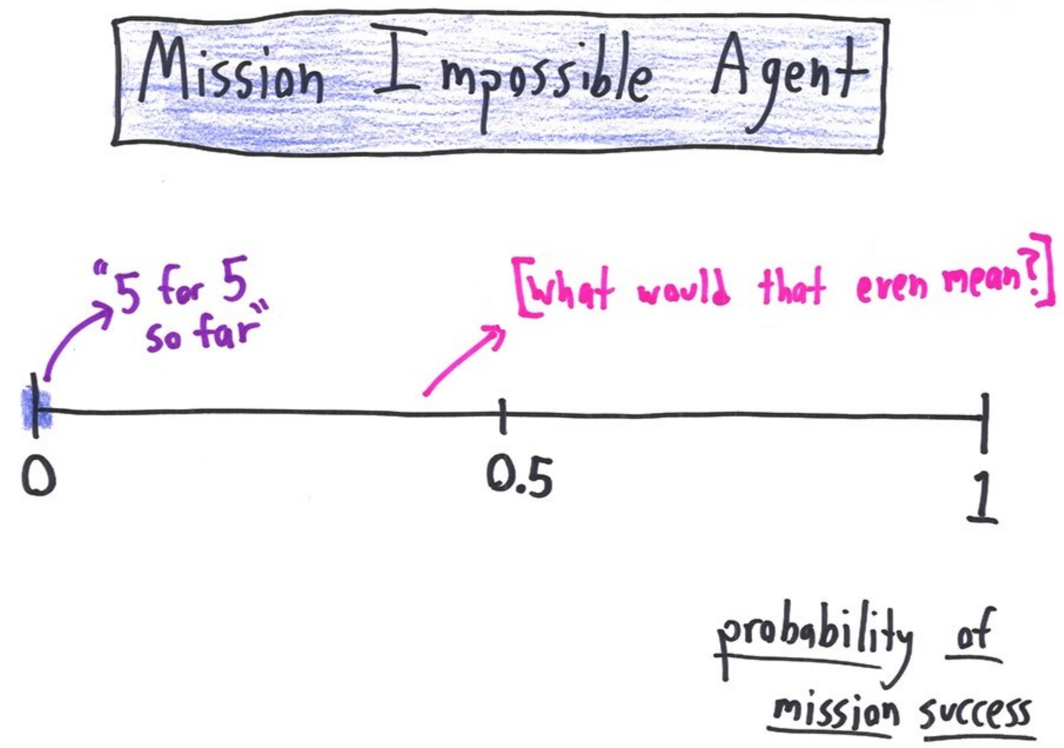
http://mathwithbaddrawings.com/2015/09/23/what-does-probability-mean-in-your-profession/

What does probability mean



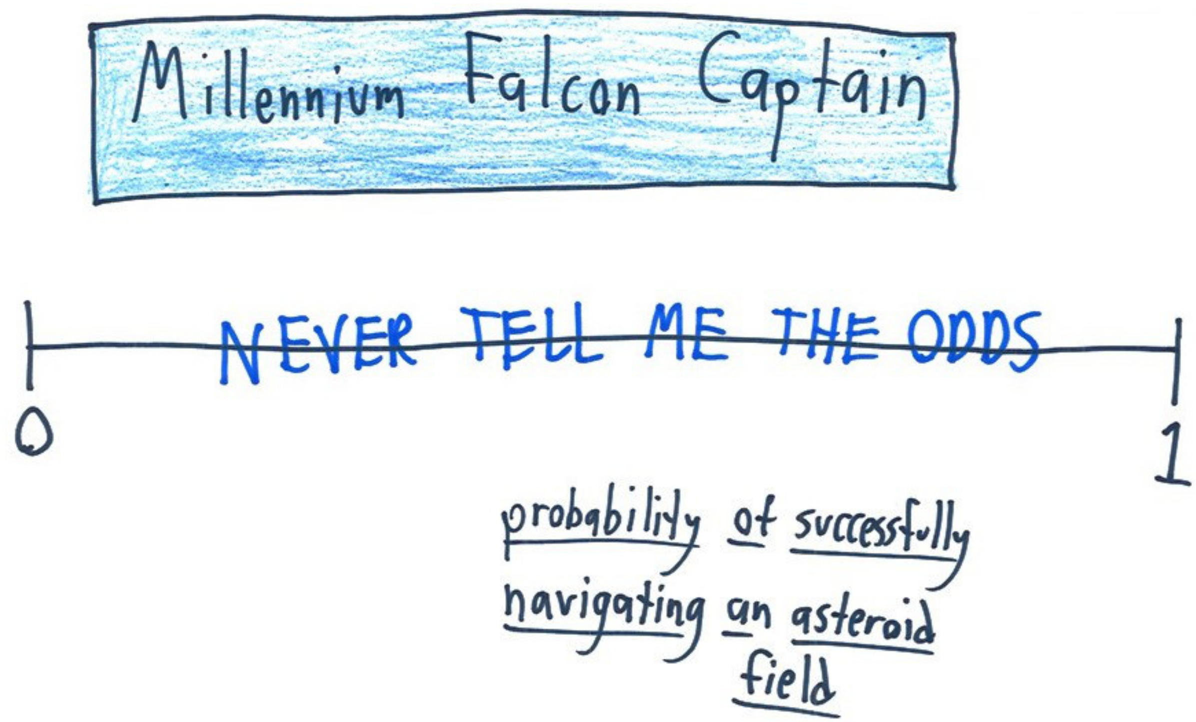
http://mathwithbaddrawings.com/2015/09/23/what-does-probability-mean-in-your-profession/

What does probability mean



http://mathwithbaddrawings.com/2015/09/23/what-does-probability-mean-in-your-profession/

What does probability mean



http://mathwithbaddrawings.com/2015/09/23/what-does-probability-mean-in-your-profession/

- Don’t confuse estimate of probability with your actual loss function.

Regression loss function

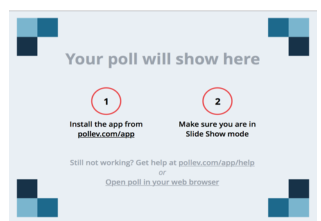
- For a linear regression predicting dollar amounts (e.g. income, housing prices) what is the residual plot likely to look like?

- Does this meet the assumptions of the linear regression model?
  - If not, how could you fix it?

Precision, Recall, Sensitivity, Specificity and ROC curves

Have you seen ROC curves?

A) Yes
B) No



Ways to be right or wrong

Claim\Is	True Yes	True No
Classify Yes	True Positive	False Positive
Classify No	False Negative	True Negative

Accuracy = (TP + TN)/(TP+FP+FN+TN)

Measuring Performance

- **Accuracy (symmetric)**
  - % correctly classified
- **Asymmetric measures**
  - **Precision**
    - P(yes | predicted as yes)
  - **Recall (or Sensitivity)**
    - P(predicted as yes | yes)
  - **Specificity**
    - P (predicted as no)| no)

Precision/Recall Sensitivity/Specificity

Claim\Is	True Yes	True No
Classify Yes	True Positive	False Positive
Classify No	False Negative	True Negative

- **Precision**
  - P(good | predicted as good) = TP/(TP+FP)
- **Recall (or Sensitivity)**
  - P(predicted as good | good) = TP/(TP+FN)
- **Specificity**
  - P (predicted as bad)| bad) = TN/(TN+FP)

Precision/Recall Example

Claim\Is	True Good	True Not Good	
Classify "Good"	70	50	
Classify "Not good"	30	350	
			500

- **Precision**
  - P(good | predicted as good) = 70/(70+50)
- **Recall (or Sensitivity) = True Positive Rate (TPR)**
  - P(predicted as good | good) = 70/(70+30)
- **Specificity = 1 – (False Positive Rate)**
  - P (predicted as bad| bad) = 350/(350+50)

F1 combines Precision+Recall

Claim\Is	True Yes	True No	
Classify Yes	True Positive	False Positive	
Classify No	False Negative	True Negative	

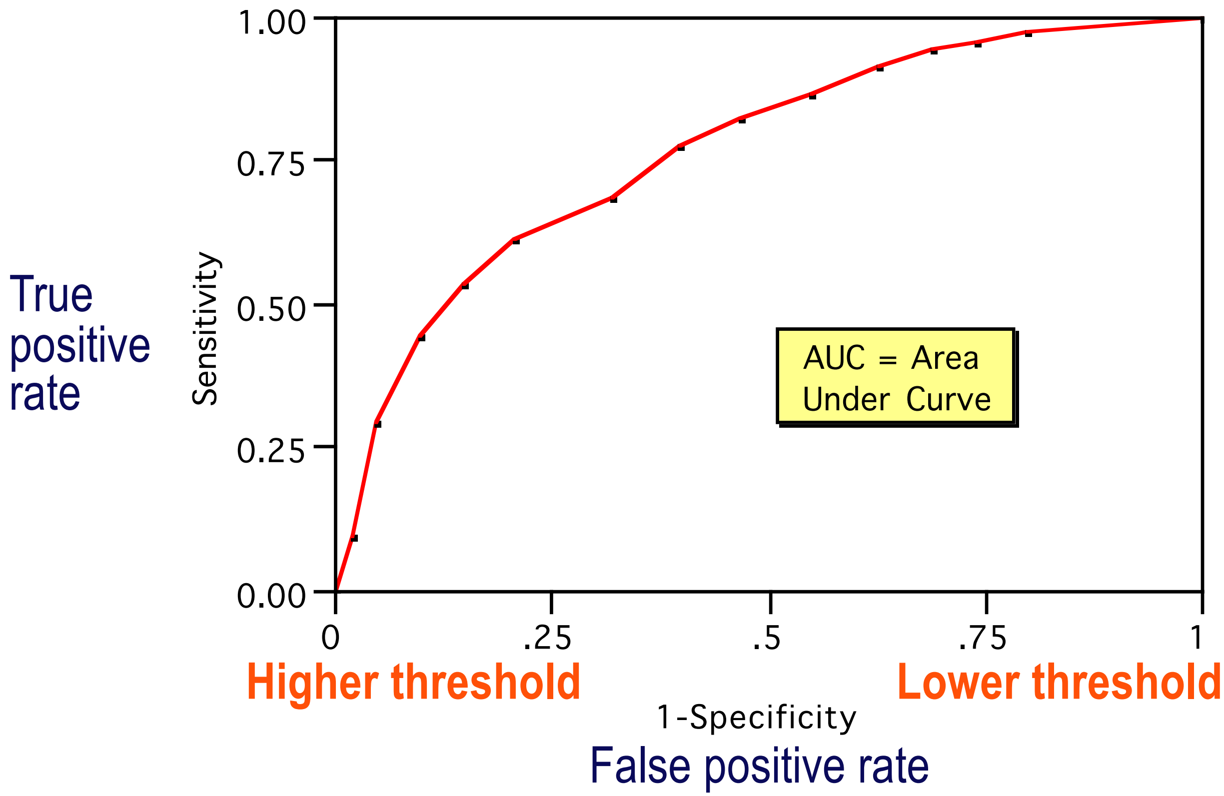
- **Precision**
  - TP/(TP+FP)
- **Recall**
  - TP/(TP+FN)
- **F1**
  - 2 precision * recall/(precision + recall)

ROC (Receiver Operarting Characteristic) Curve

- Sort all examples from highest probability (or score) of being ‘yes’, p(y=‘yes’|**x**), to lowest
- Sweep the threshold for predicting an example to be labeled ‘yes’ from 1 down to 0
  - This varies *specificity* from 1 to 0.
- At each threshold compute the *sensitivity*
  - i.e., the fraction of the true positives you found
- Plot the curve

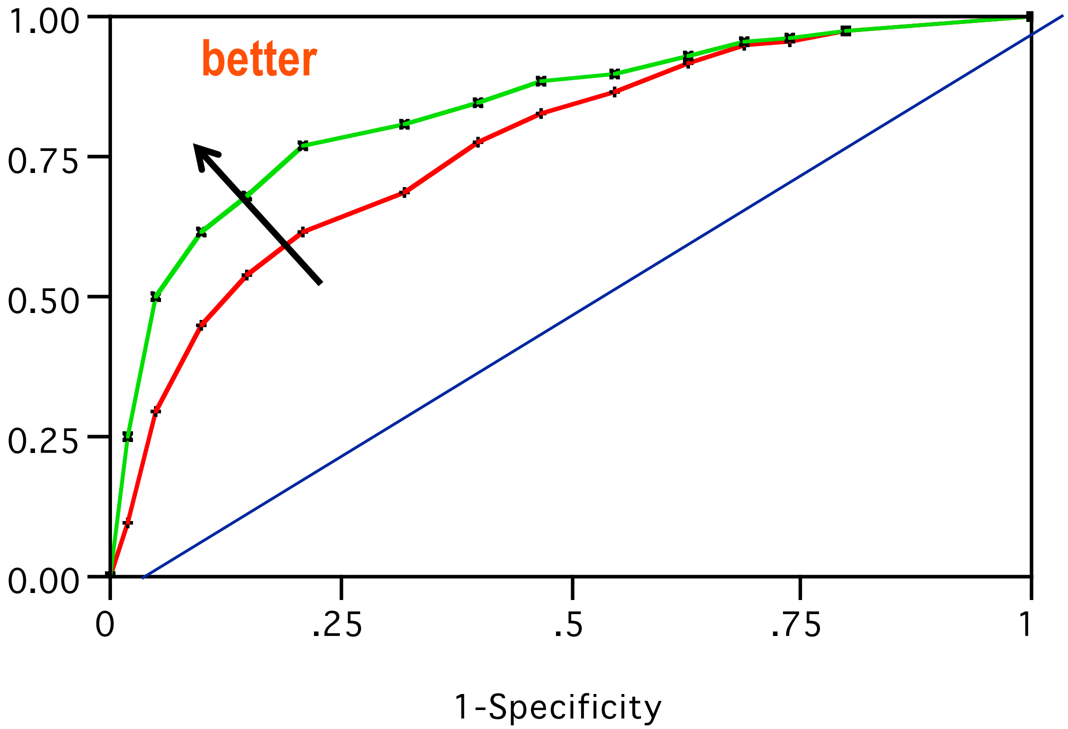
https://en.wikipedia.org/wiki/Receiver_operating_characteristic

ROC Chart Varies Threshold



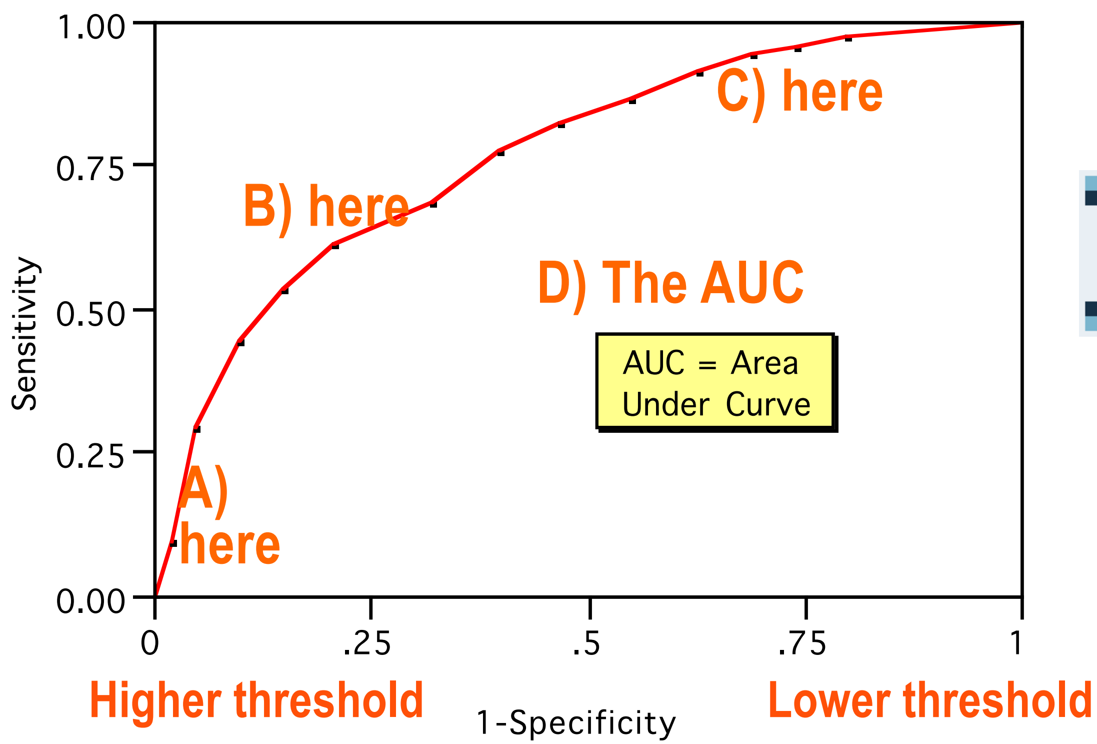
p(y|x) > threshold **to be in class**

ROC charts support comparison



AUC = 0.5 is random guessing
AUC = 1.0 is perfection

Where does google care about?

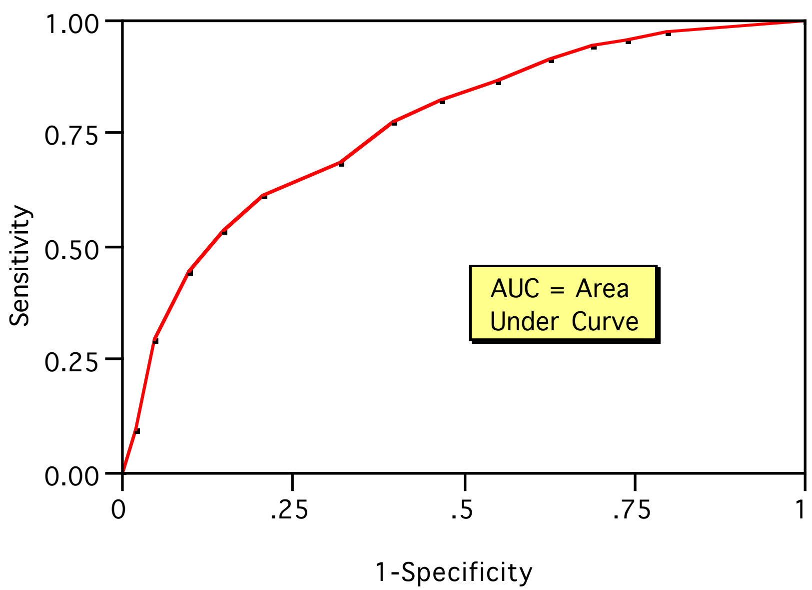


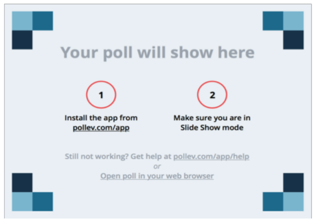
p(y|x) > *threshold* to be in class

- **Which method is most likely to be better for generating an ROC curve?**

  A) Logistic regression

  B) SVM





Confusion Matrix

- A confusion matrix shows the counts of the actual versus predicted class values.
- Example (overall accuracy rate of 73.9%)

		Actual Class		
		Class A	Class B	Class C
Predicted	Class A	20	5	2
Class	Class B	6	20	4
	Class C	4	2	25

For the confusion matrix

		Actual	
		purchase	no purchase
Predicted	purchase	10	60
	no purchase	20	200

- What is its precision?
- What is its recall?
- How do you
  a) increase precision (but decrease recall)
  b) increase both precision and recall

a) 10/20
b) 10/(10+20)
c) 10/60
d) 10/(10+60
e) other