Classification Metrics

Week 05 - Day 02

Logistic Regression Summary

1 / (1 + e**-(b0 + b1*x1 + … + bn*xn))

f(features) = probability, [0,1]

(using logistic function)

g(probability) -> {0,1}

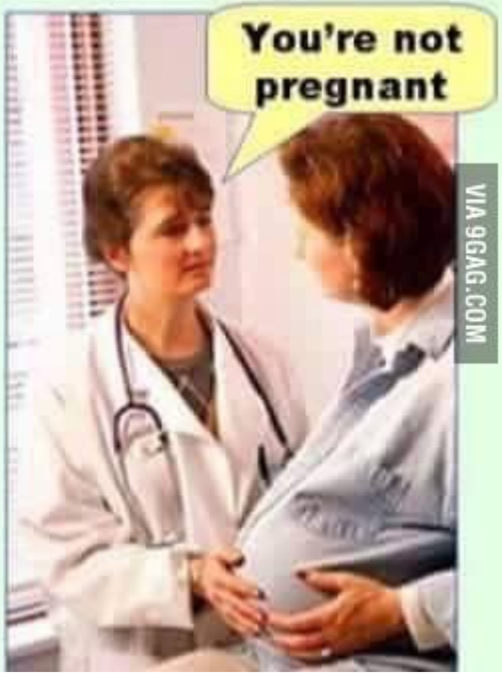
(using cutoff, e.g 0.5)

Confusion Matrix

Let’s focus on binary classification

Type I error
(false positive)

Type II error
(false negative)



PREDICTED

REAL

	Negative	Positive
Negative	40	10
Positive	5	45

		PREDICTED	
		Negative	Positive
REAL	Negative	TN	FP
	Positive	FN	TP

Cutoffs and Predictions

Different cutoffs → different predictions

P			
0.3			
0.45			
0.55			
0.65			
0.72			

P	Prediction w/cutoff 0.5		
0.3	negative		
0.45	negative		
0.55	positive		
0.65	positive		
0.72	positive		

P	Prediction w/cutoff 0.5	Prediction w/cutoff 0.4	
0.3	negative	negative	
0.45	negative	positive	
0.55	positive	positive	
0.65	positive	positive	
0.72	positive	positive	

P	Prediction w/cutoff 0.5	Prediction w/cutoff 0.4	Prediction w/cutoff 0.6
0.3	negative	negative	negative
0.45	negative	positive	negative
0.55	positive	positive	negative
0.65	positive	positive	positive
0.72	positive	positive	positive

Precision and Recall

Information Retrieval

&

Machine Learning

PREDICTED

REAL	Neg	Pos
Neg	850	50
Pos	10	990

Recall

% of positive real values I can “recall”

(i.e. correctly predict as positive)

<u>Query + Corpus of documents</u>

How many documents can I correctly retrieve?

Recall = TP/(FN+TP) = TP / REAL_POS_VALUES

PREDICTED

		Neg	Pos
REAL	Neg	TN	FP
	Pos	FN	TP

Precision

How many positive test are really positive?

Query + Corpus of documents

How many documents are really "interesting", out of the ones I'm showing?

Precision = TP/(FP+TP) = TP / POSITIVE_TESTS

PREDICTED

REAL		Neg	Pos
	Neg	TN	FP
	Pos	FN	TP

Recall and Precision are percentages

([0, 1], [0%, 100%])

Is recall==100% good?

PREDICTED

		Negative	Positive
REAL	Negative	0	100
	Positive	0	100

Is precision==100% good?

PREDICTED

REAL		Negative	Positive
	Negative	100	0
	Positive	99	1

Precision
vs.
Recall

Different cutoffs

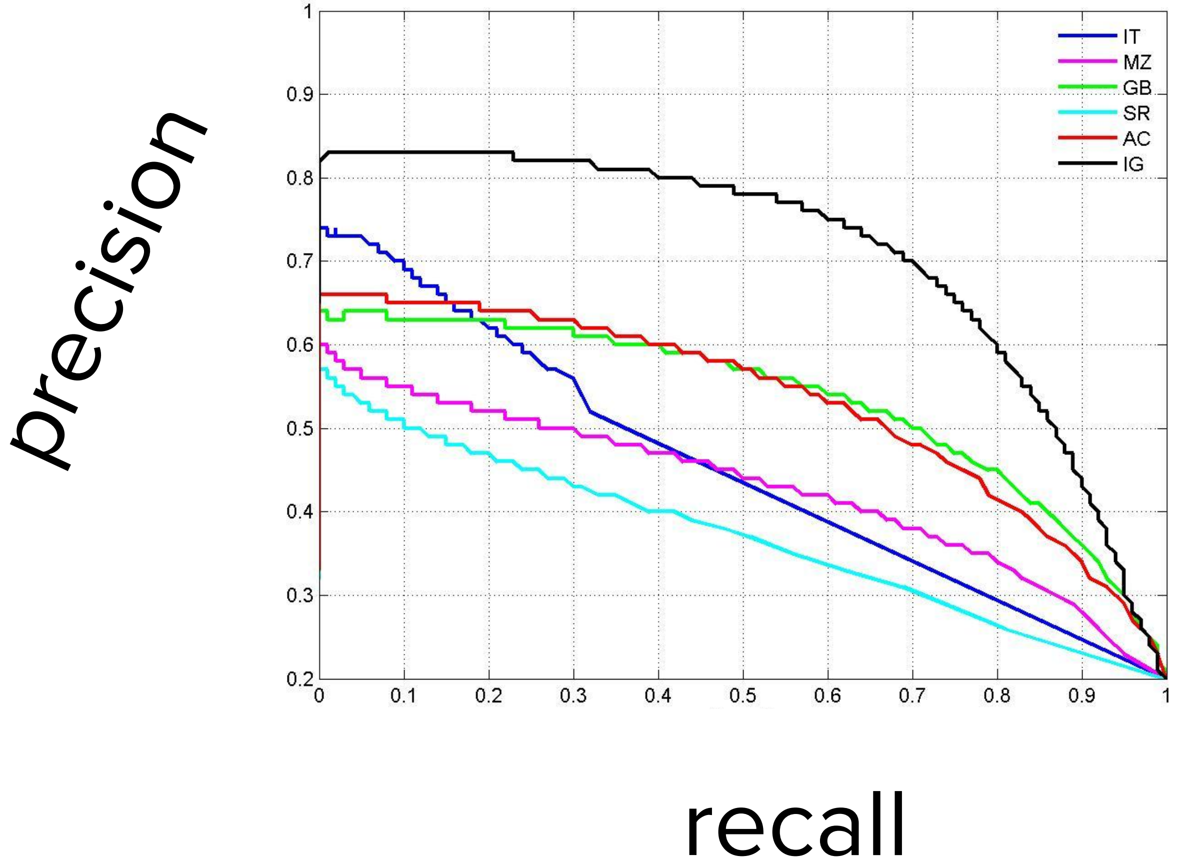
=

Different predictions

=

Different precisions & recalls

Models comparison



Precision/Recall = Yin & Yang

Finding the “best” precision/recall

=

Business decision

Disease: which one is better?

PREDICTED

REAL	Neg	Pos
Neg	800	200
Pos	200	800

PREDICTED

REAL	Neg	Pos
Neg	998	2
Pos	400	600

Giving a loan: which one is better?

PREDICTED

REAL	Neg	Pos
Neg	800	200
Pos	200	800

PREDICTED

	Neg	Pos
Neg	998	2
Pos	400	600

$$\text{F1-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Specificity and Sensitivity

Used in Medicine

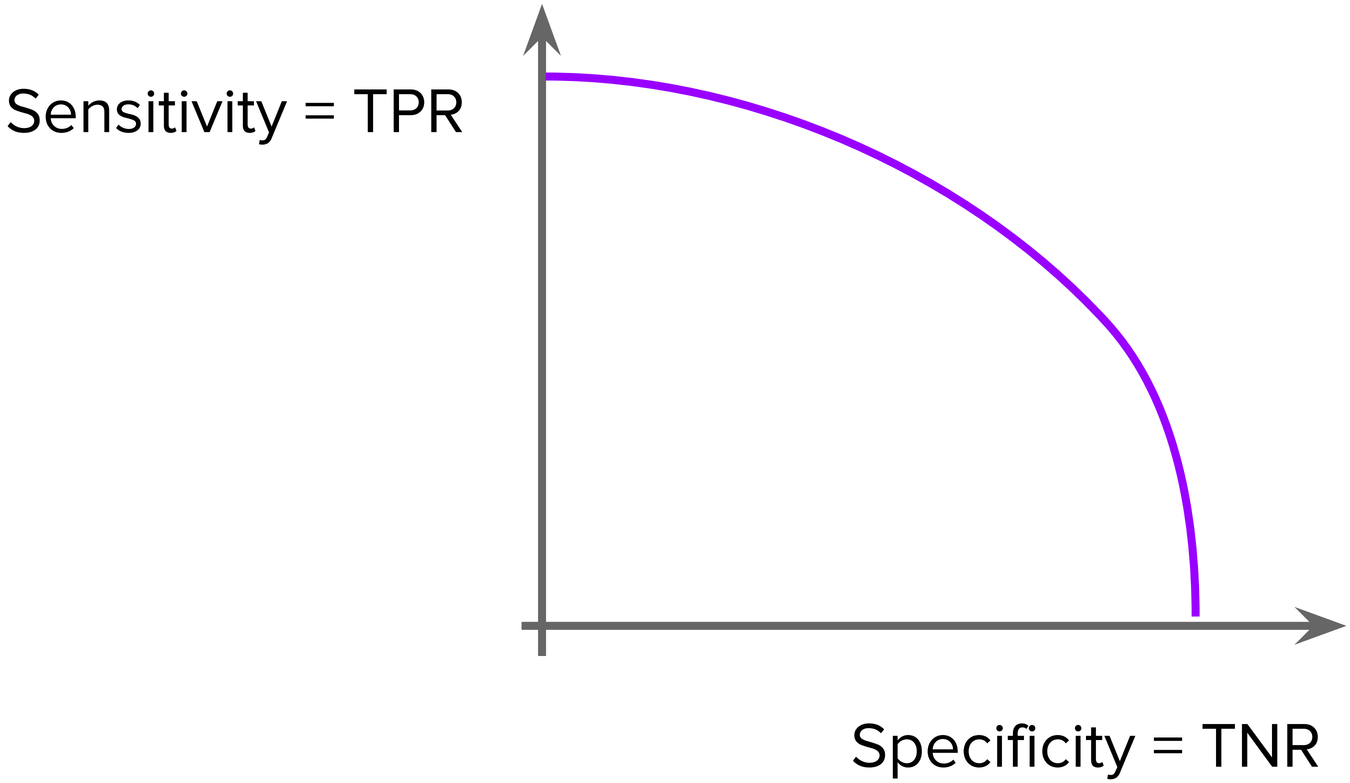
- Sensitivity
  - Recall!
  - True positive rate
  - TP / REAL_POSITIVE
- Specificity:
  - True negative rate
  - TN / REAL_NEGATIVE

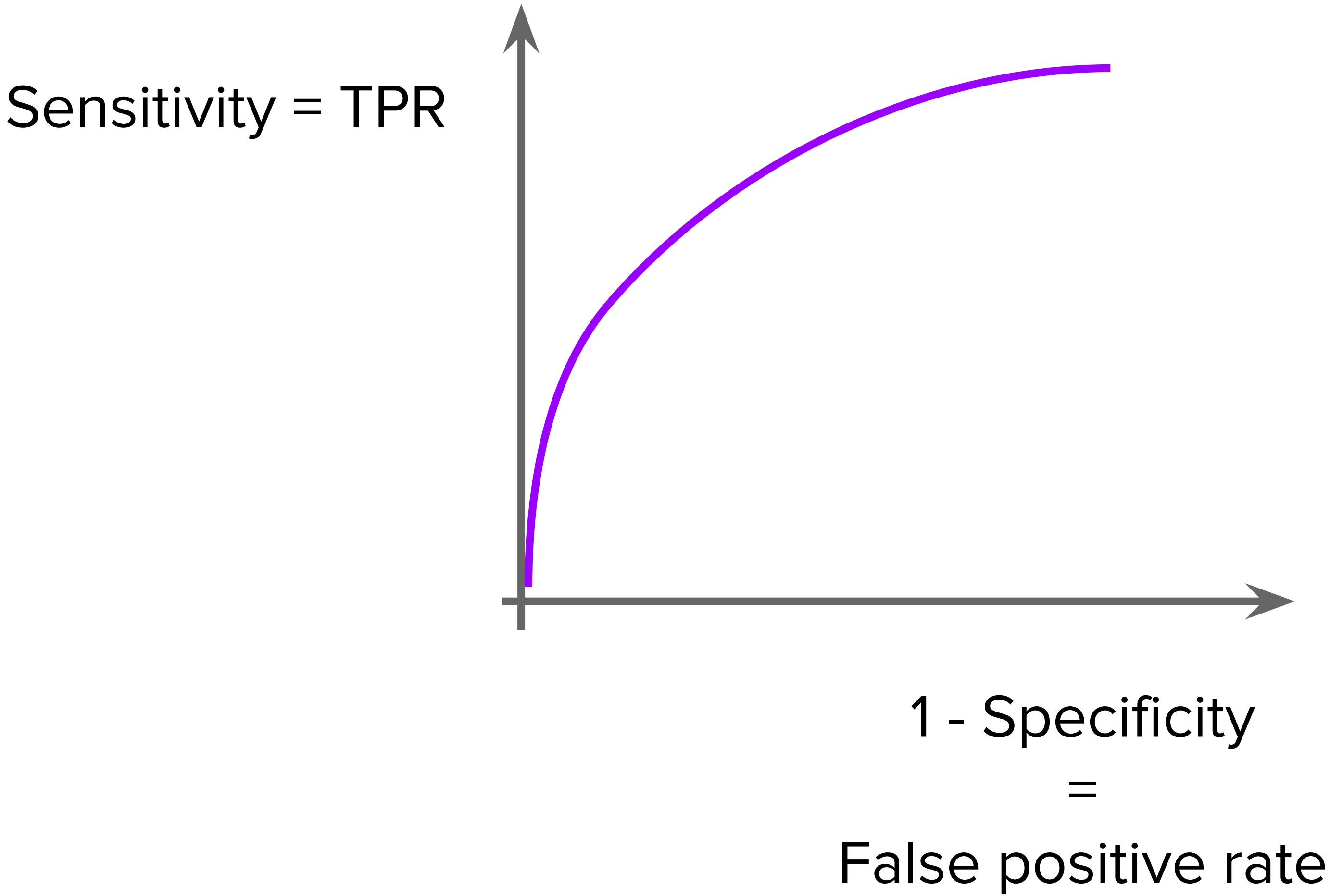
Specificity = TN/(TN+FP) = TN / REAL_NEGATIVE

PREDICTED

REAL

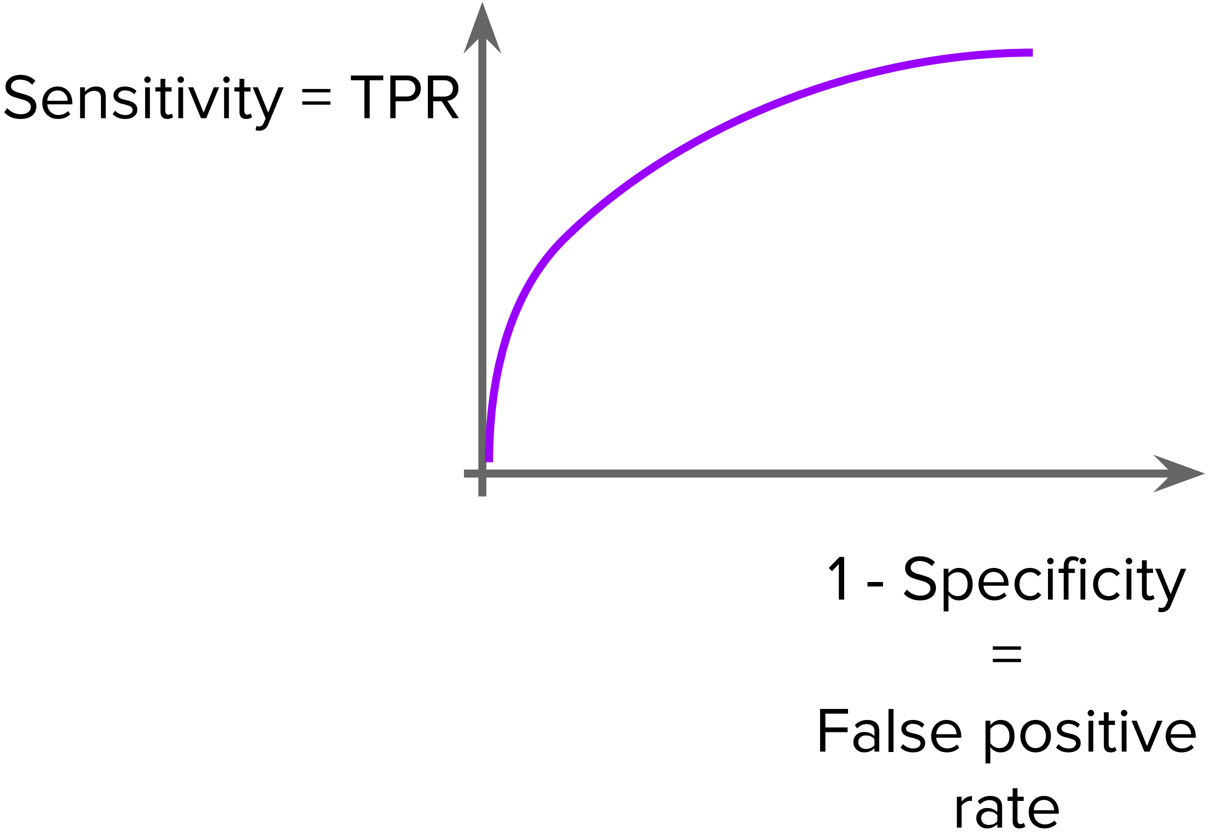
	Neg	Pos
Neg	TN	FP
Pos	FN	TP


Sensitivity = TPR
Specificity = TNR

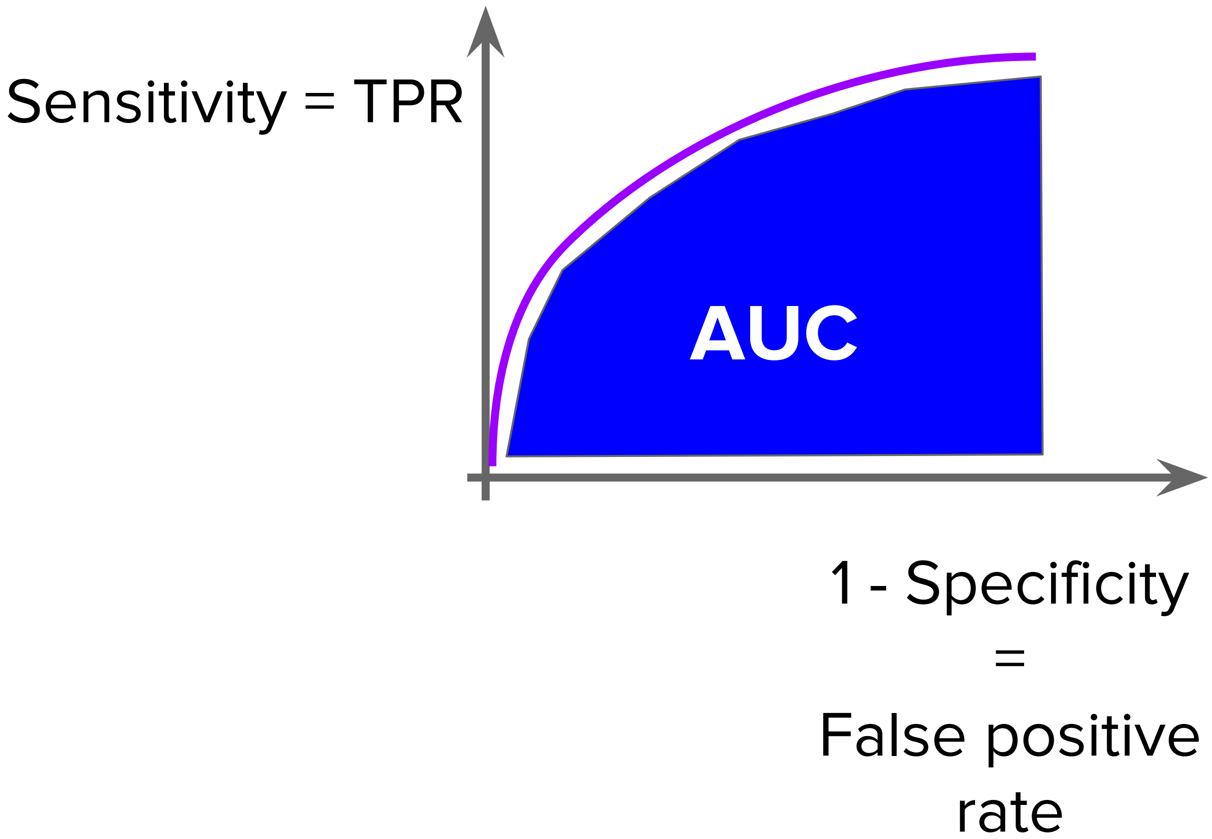

Sensitivity = TPR
1 - Specificity
=
False positive rate

ROC and AUC

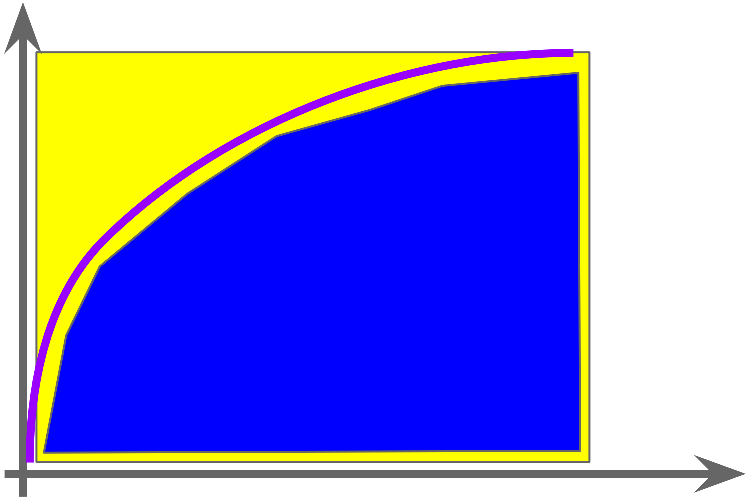
ROC CURVE

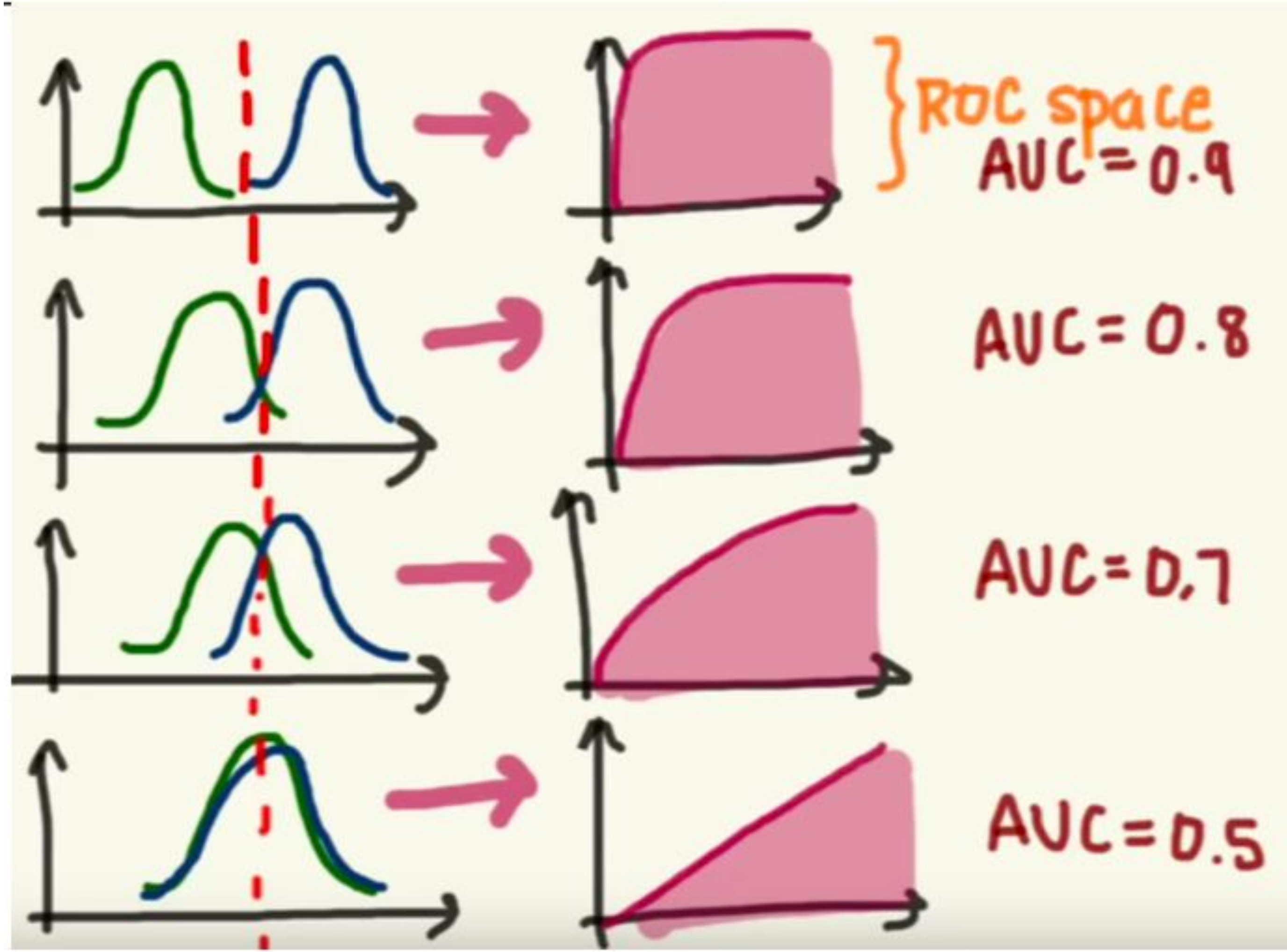


Area Under the (Roc) Curve = AUC



AUC = AUC (blue) / ALL THE AREA (yellow)





ROC space
AUC = 0.9
AUC = 0.8
AUC = 0,7
AUC = 0.5

“Cutoff - agnostic” Metrics

P	Prediction w/cutoff 0.5	Prediction w/cutoff 0.4	Prediction w/cutoff 0.6
0.3			
0.45			
0.55			
0.65			
0.72			

- Related to a single cutoff
    - Precision / recall
    - Specificity / sensitivity

- “Cutoff-Agnostic”:
    - AUC

AUC works even if your dataset is unbalanced!

Summary

- Different cutoffs = different predictions

- Different cutoffs = different predictions
- Best cutoff = business problem

- Different cutoffs = different predictions
- Best cutoff = business problem
- Precision/Recall - ML

- Different cutoffs = different predictions
- Best cutoff = business problem
- Precision/Recall - ML
- Sensitivity/specificity - Medicine

- Different cutoffs = different predictions
- Best cutoff = business problem
- Precision/Recall - ML
- Sensitivity/specificity - Medicine
- Roc curve

- Different cutoffs = different predictions
- Best cutoff = business problem
- Precision/Recall - ML
- Sensitivity/specificity - Medicine
- Roc curve
- AUC - cutoff-agnostic and good for unbalanced datasets