

# Evaluating Performance I

Lecture 06

# Choose 2

Good

Cheap

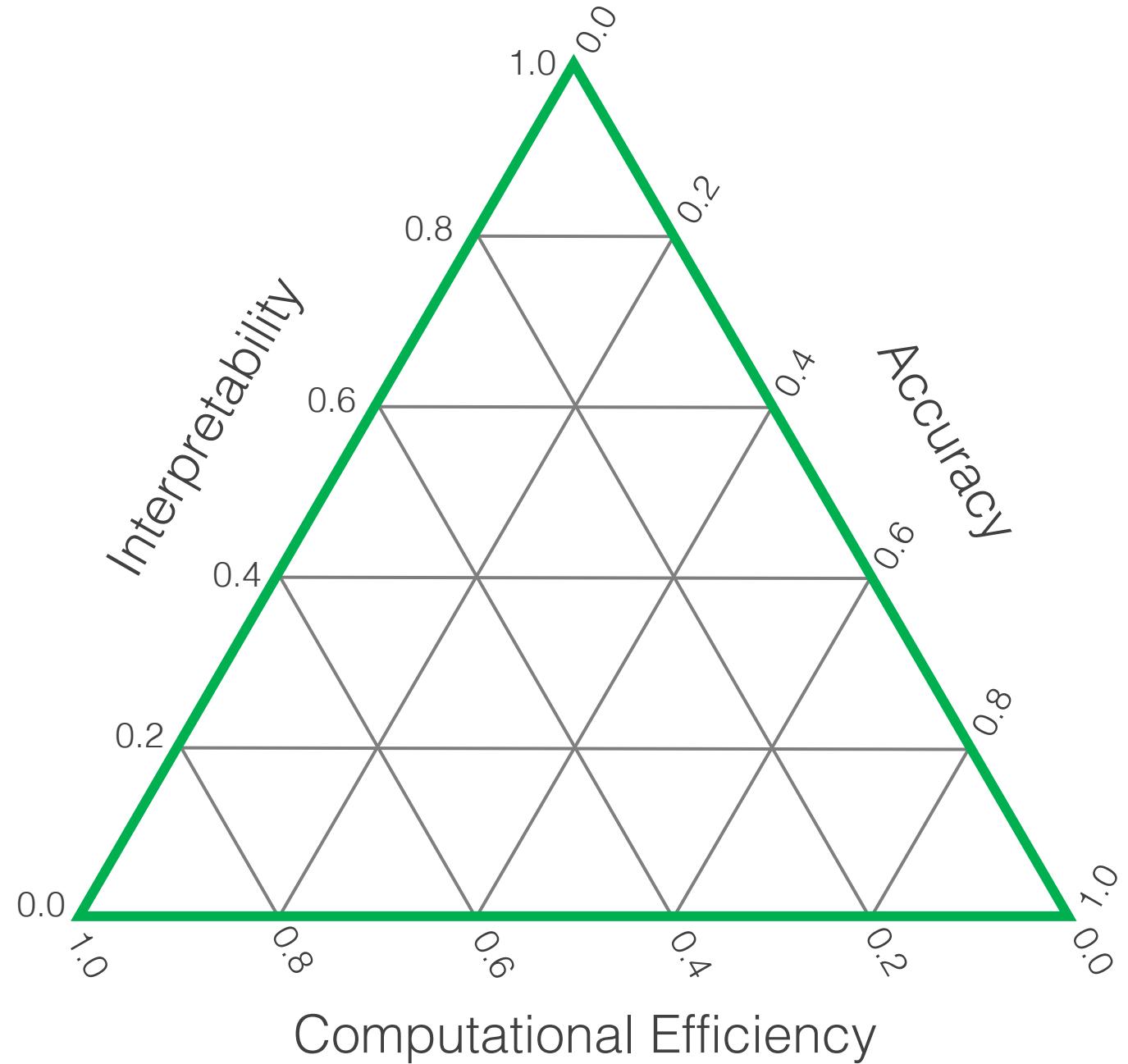
Fast

# Modeling Tradeoffs

Interpretability

Computational Efficiency

Accuracy



# Supervised Learning Performance Evaluation

## Regression

- Mean squared error (MSE)
- Mean absolute error (MAE)
- $R^2$ , coefficient of determination
- Adjusted  $R^2$
- Explained variance

## Classification

### Binary

Receiver Operating  
Characteristic (ROC)  
curves

### Multiclass

Confusion matrices

## Common Metrics

---

- Classification accuracy
- True positive rate
- False positive rate
- Precision
- $F_1$  Score
- Area under the ROC curve  
(AUC)

- Classification accuracy
- Micro-averaged  $F_1$  Score
- Macro-averaged  $F_1$  Score

# Regression: Mean Squared Error

The mean squared error (MSE)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Absolute measure of performance

One of the most widely used loss / cost functions

# Regression: Mean **Absolute** Error

The mean absolute error (MAE)

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Absolute measure of performance

# $R^2$ Coefficient of determination

Proportion of the response variable variation explained by the model

Residual sum of squares  
(variation in the residuals)

$$SS_{res} = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Total sum of squares  
(variation in the data)

$$SS_{tot} = \sum_{i=1}^N (y_i - \bar{y})^2$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

R-squared

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Relative measure of performance

# R<sup>2</sup> Coefficient of determination

R<sup>2</sup> increases with more predictor variables

Adjusted R squared:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{N - 1}{N - p - 1}$$

Adjusts R squared to account for the number of predictor variables

This value is always less than or equal to the unadjusted R squared

# Explained Variance

Proportion of the response variable variation explained by the model

$$\text{Explained Variance} = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)}$$

$$\begin{aligned}\text{Var}(y - \hat{y}) &= \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 - \left( \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i) \right)^2 \\ &= \frac{1}{N} SS_{res} - (\text{mean error})^2\end{aligned}$$

$$\text{Var}(y) = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

This will equal  $R^2$  when the mean of  $y - \hat{y}$  is zero

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

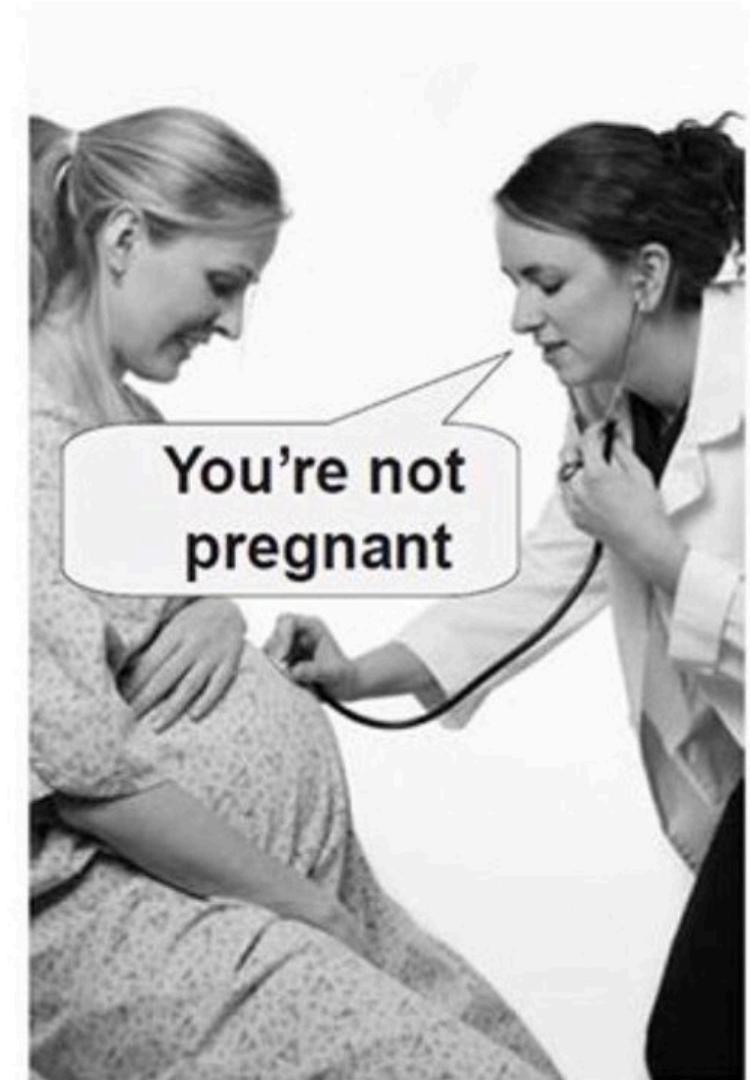
Relative measure of performance

# Types of error

**False Positive**  
(Type I error)



**False Negative**  
(Type II error)



# Binary Classification

		Predicted Class, $\hat{y}$	
		Class 1 (target)	Class 0 (non-target)
True Class, $y$	Class 1 (target)	true positive	false negative
	Class 0 (non-target)	false positive	true negative

Type II Error  
(missed target)

Type I Error

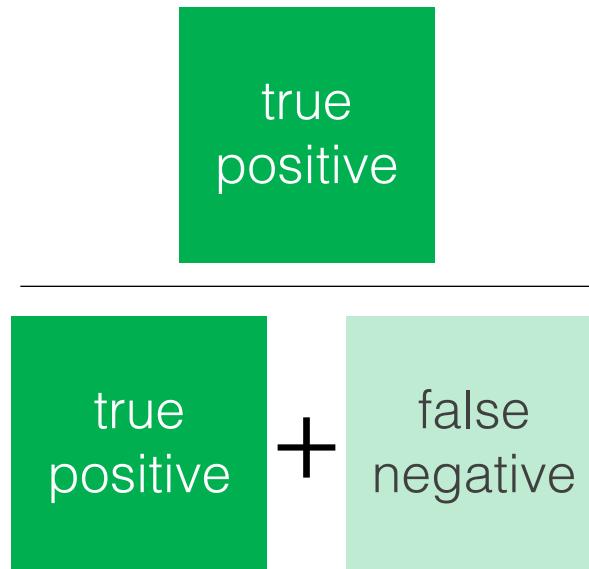
# Binary Classification

Predicted Class,  $\hat{y}$

		Class 1 (target)	Class 0 (non-target)
True Class, $y$	Class 1 (target)	true positive	false negative
	Class 0 (non-target)	false positive	true negative

True positive rate  
Probability of detection,  $p_D$

Sensitivity  
Recall



How many targets (Class 1)  
were correctly classified as  
targets?

# Binary Classification

Predicted Class,  $\hat{y}$

		Class 1 (target)	Class 0 (non-target)
		true positive	false negative
True Class, $y$	Class 1 (target)		
	Class 0 (non-target)	false positive	true negative

False positive rate

Probability of false alarm,  $p_{FA}$

$$\frac{\text{false positive}}{\text{false positive} + \text{true negative}}$$

How many non-targets (Class 0) were incorrectly classified as targets?

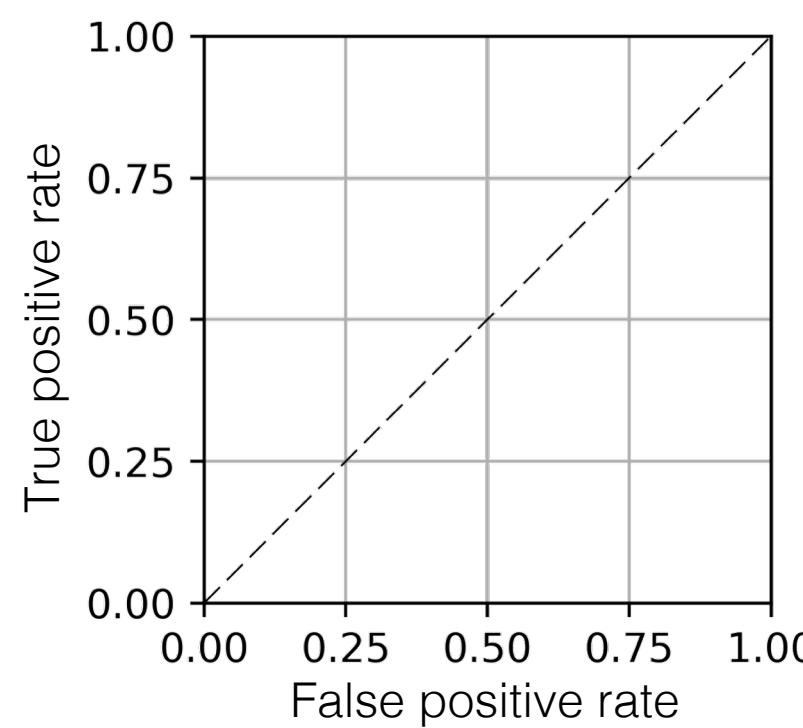
# Binary Classification

Predicted Class,  $\hat{y}$

		Class 1 (target)	Class 0 (non-target)	Precision
		true positive	false negative	<hr/>
True Class, $y$	Class 1 (target)	<hr/>		<hr/>
	Class 0 (non-target)	false positive	true negative	<hr/>

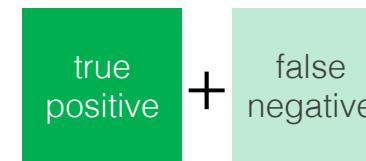
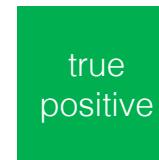
How many of the predicted targets are targets?

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



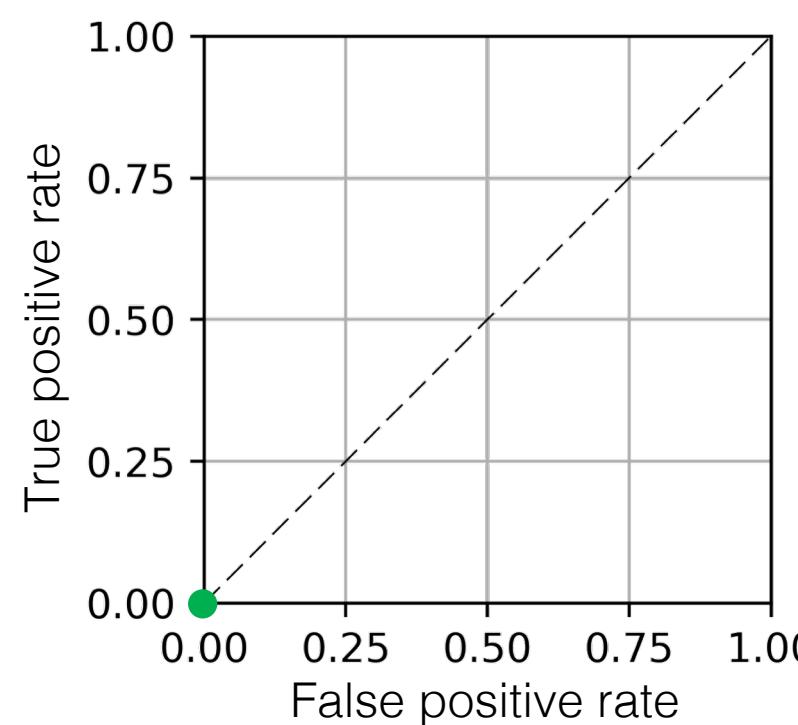
Total Positives = 3

Total Negatives = 2

Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
-----------	------------------	--------------------	-------------------	---------------------

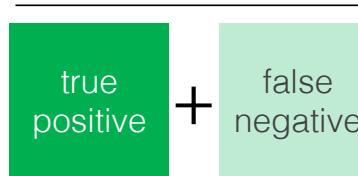
True Class Label (y)	Classifier Confidence
1	1.40
1	0.95
0	0.80
1	0.60
0	-0.10

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



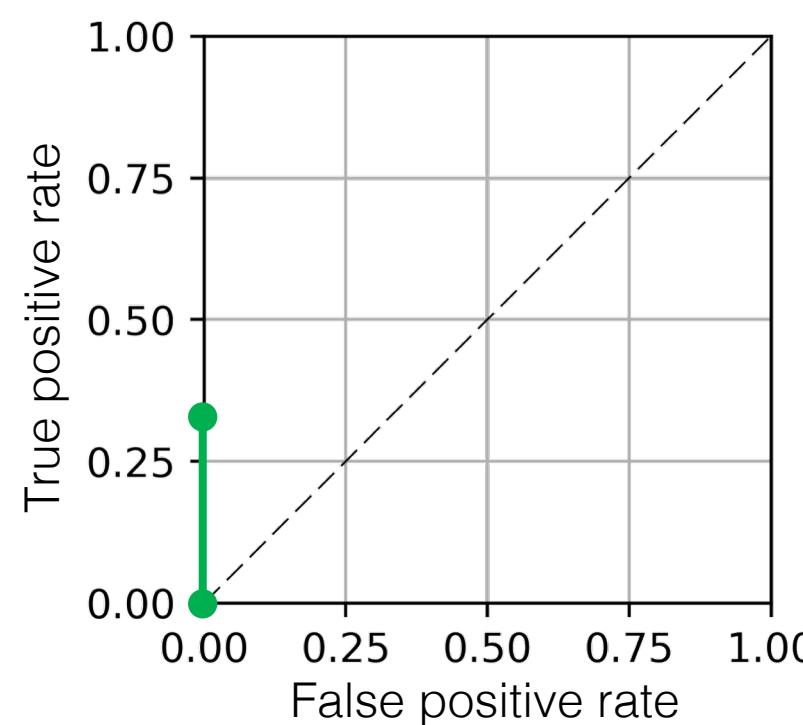
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
0	1	1.40
0	1	0.95
0	0	0.80
0	1	0.60
0	0	-0.10

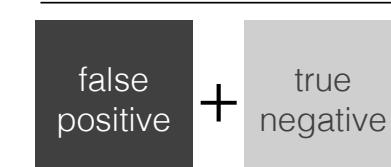
Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
$\infty$	0	0	0	0

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



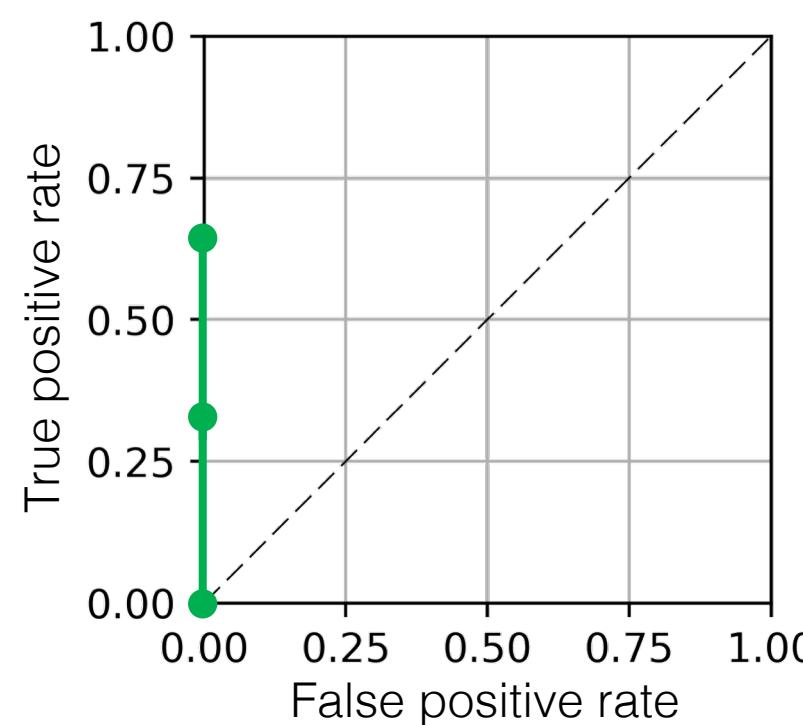
Total Positives = 3

Total Negatives = 2

Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
$\infty$	0	0	0	0
1.0	1	0.333	0	0

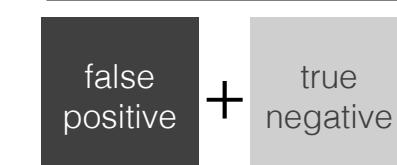
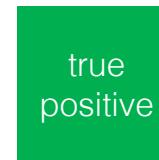
Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
0	1	0.95
0	0	0.80
0	1	0.60
0	0	-0.10

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



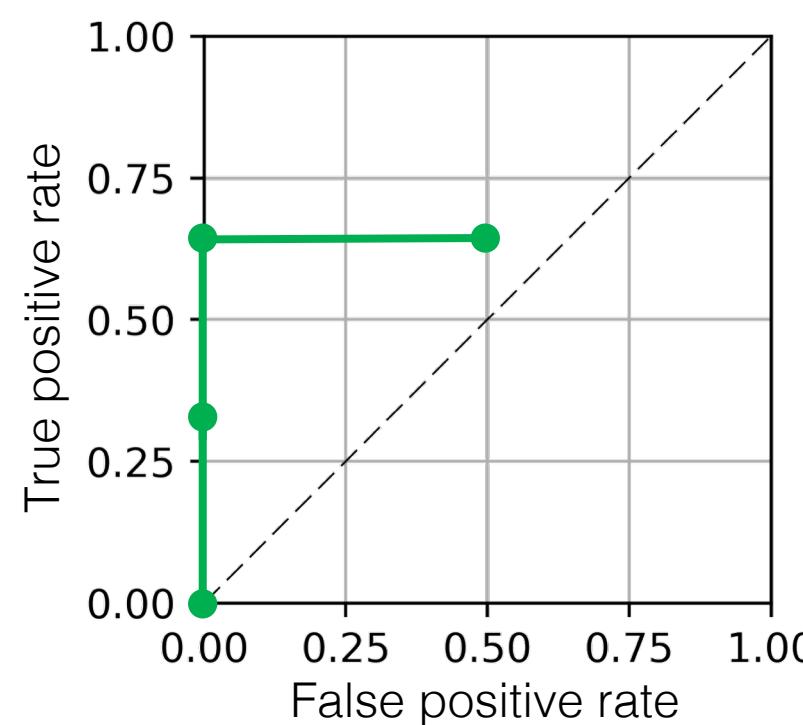
Total Positives = 3

Total Negatives = 2

Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
$\infty$	0	0	0	0
1.0	1	0.333	0	0
0.9	2	0.667	0	0

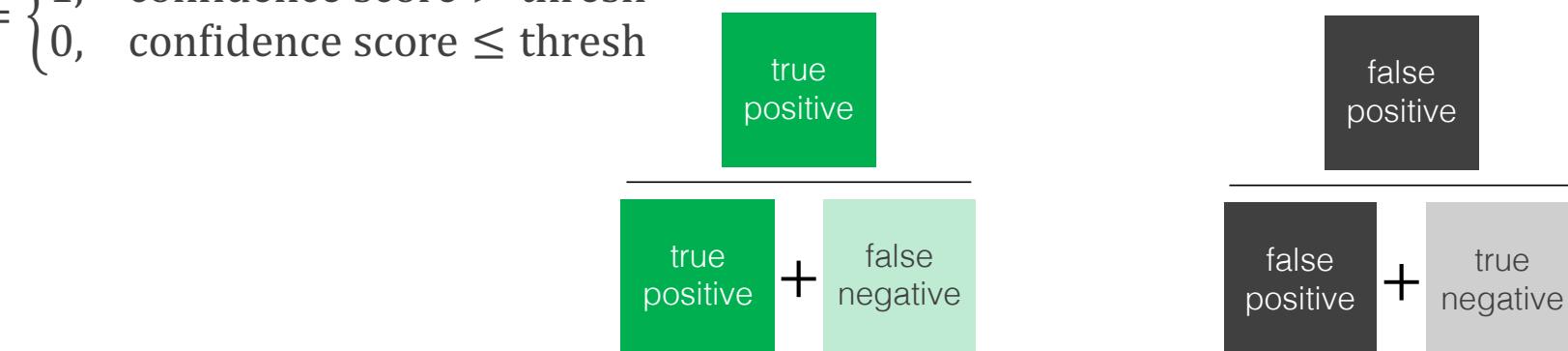
Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
0	0	0.80
0	1	0.60
0	0	-0.10

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



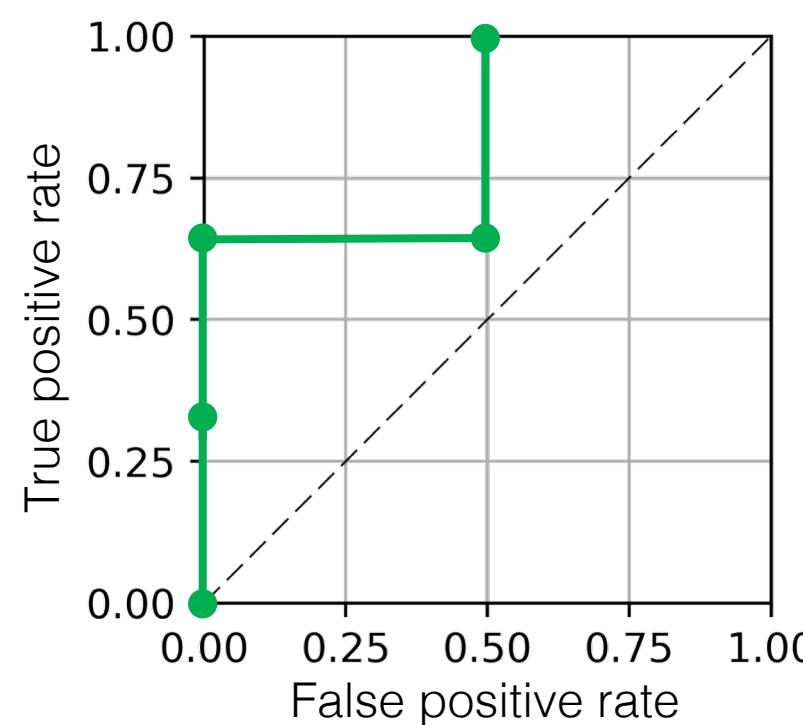
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
1	0	0.80
0	1	0.60
0	0	-0.10

Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
$\infty$	0	0	0	0
1.0	1	0.333	0	0
0.9	2	0.667	0	0
0.7	2	0.667	1	0.5

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



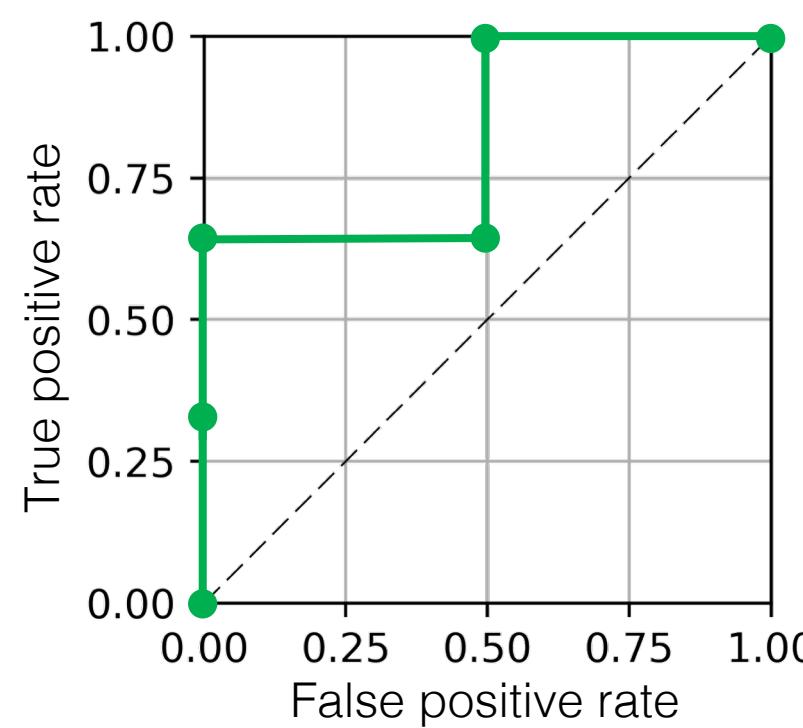
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
1	0	0.80
1	1	0.60
0	0	-0.10

Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
$\infty$	0	0	0	0
1.0	1	0.333	0	0
0.9	2	0.667	0	0
0.7	2	0.667	1	0.5
0.0	3	1	1	0.5

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



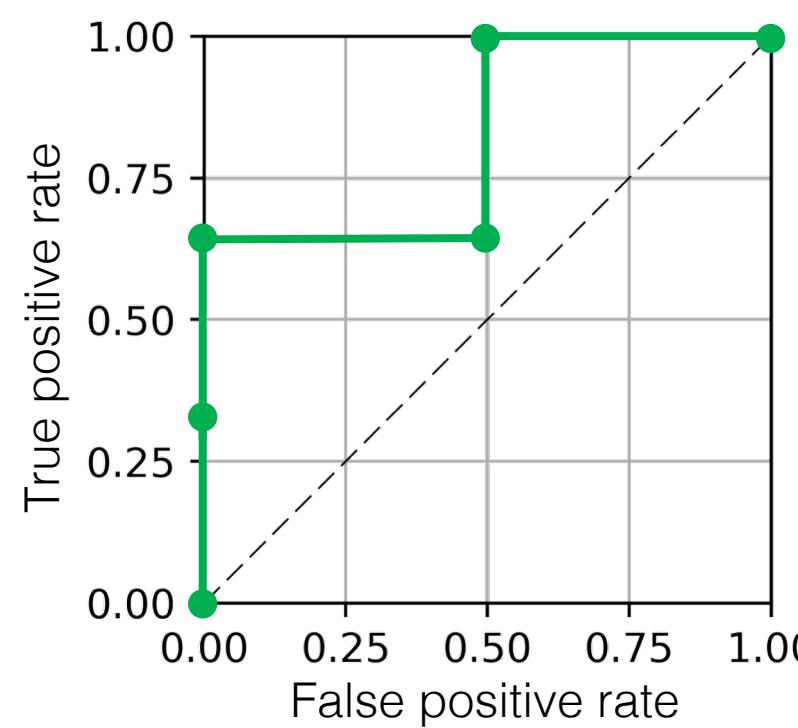
Total Positives = 3

Total Negatives = 2

Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
$\infty$	0	0	0	0
1.0	1	0.333	0	0
0.9	2	0.667	0	0
0.7	2	0.667	1	0.5
0.0	3	1	1	0.5
$-\infty$	3	1	2	1

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
1	0	0.80
1	1	0.60
1	0	-0.10

# ROC Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$

$$AUC = \left(\frac{2}{3}\right)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{2}\right) = \frac{5}{6} \cong 0.833$$

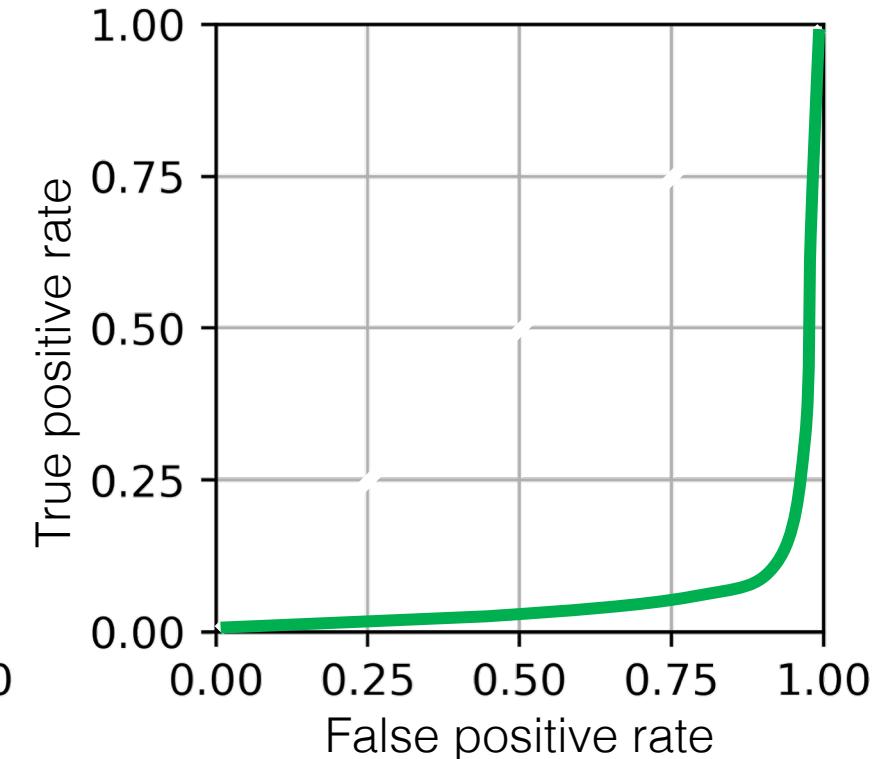
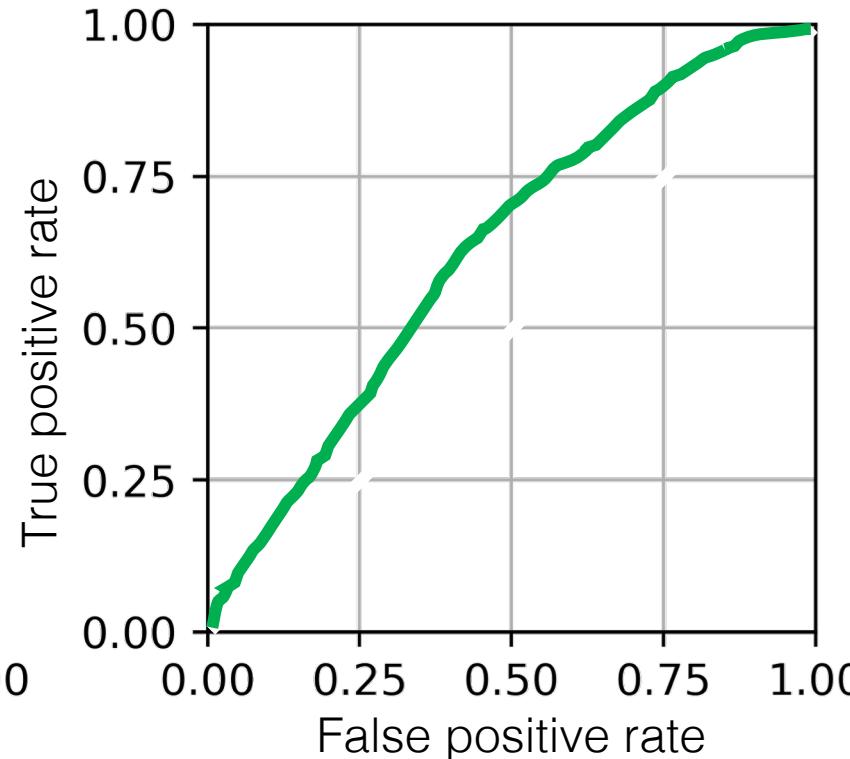
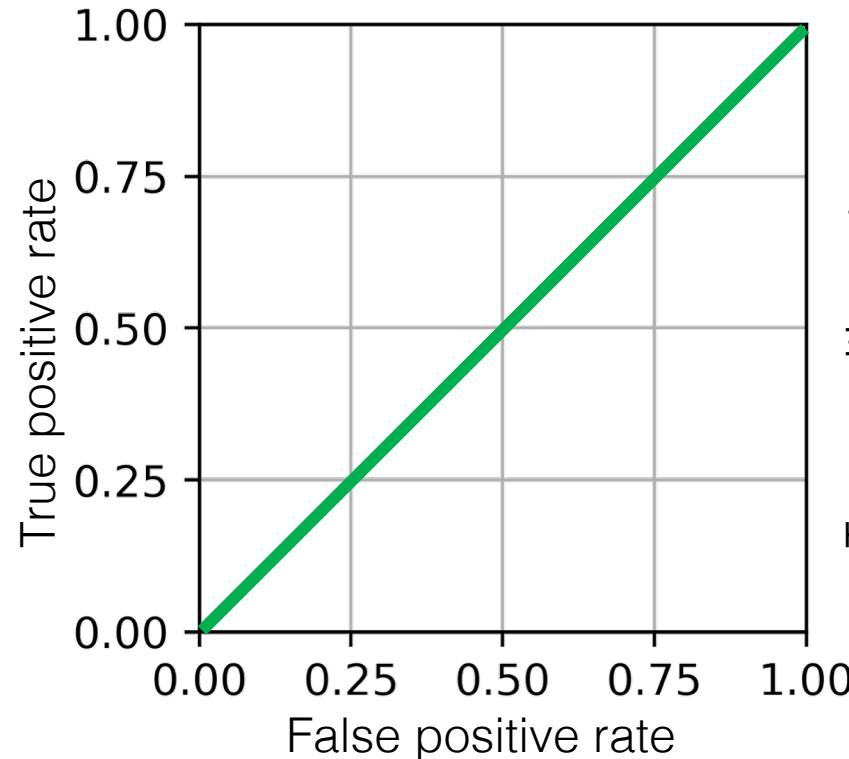
Total Positives = 3

Total Negatives = 2

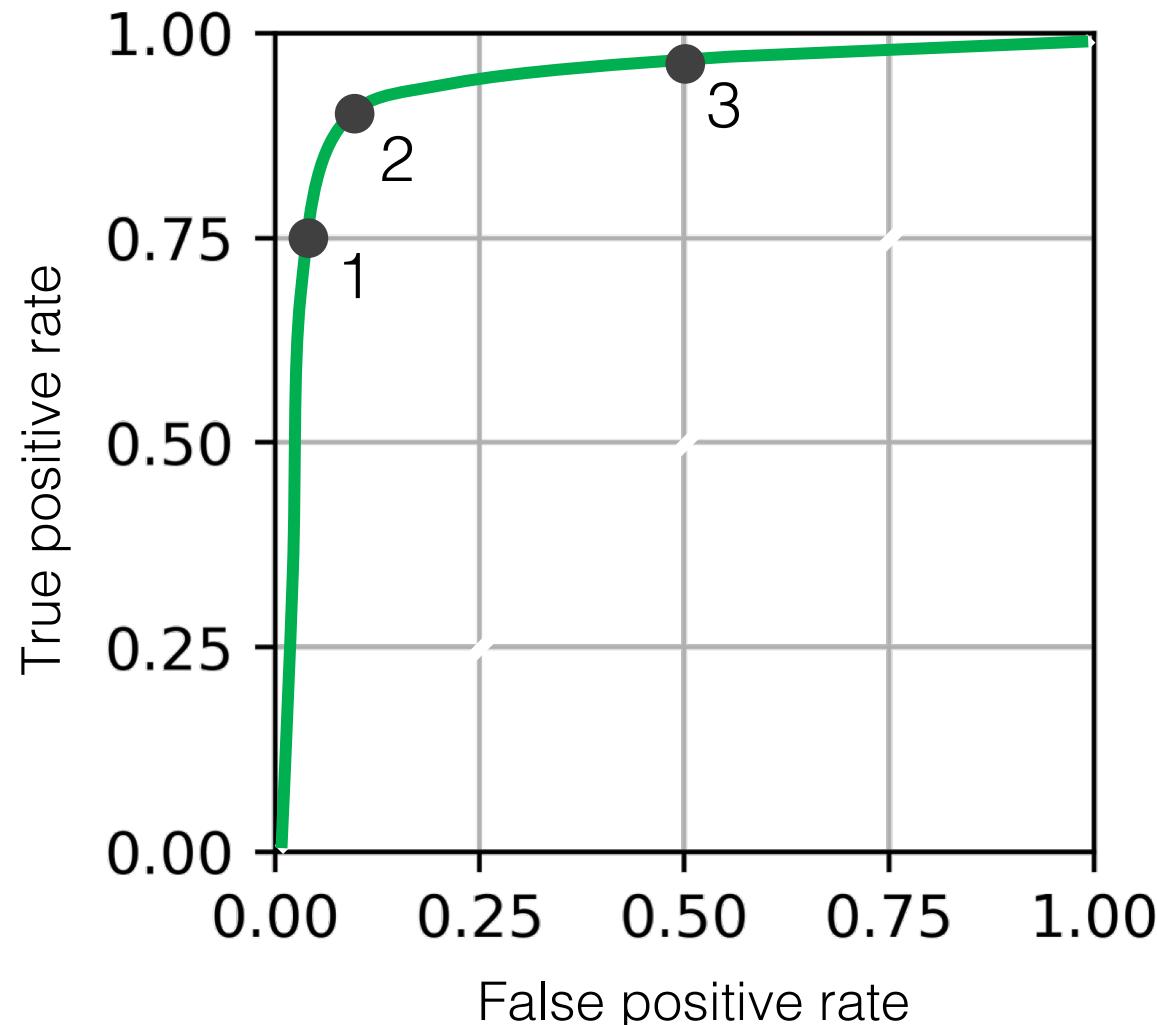
Threshold	# True Positives	True Positive Rate	# False Positives	False Positive Rate
$\infty$	0	0	0	0
1.0	1	0.333	0	0
0.9	2	0.667	0	0
0.7	2	0.667	1	0.5
0.0	3	1	1	0.5
$-\infty$	3	1	2	1

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
1	0	0.80
1	1	0.60
1	0	-0.10

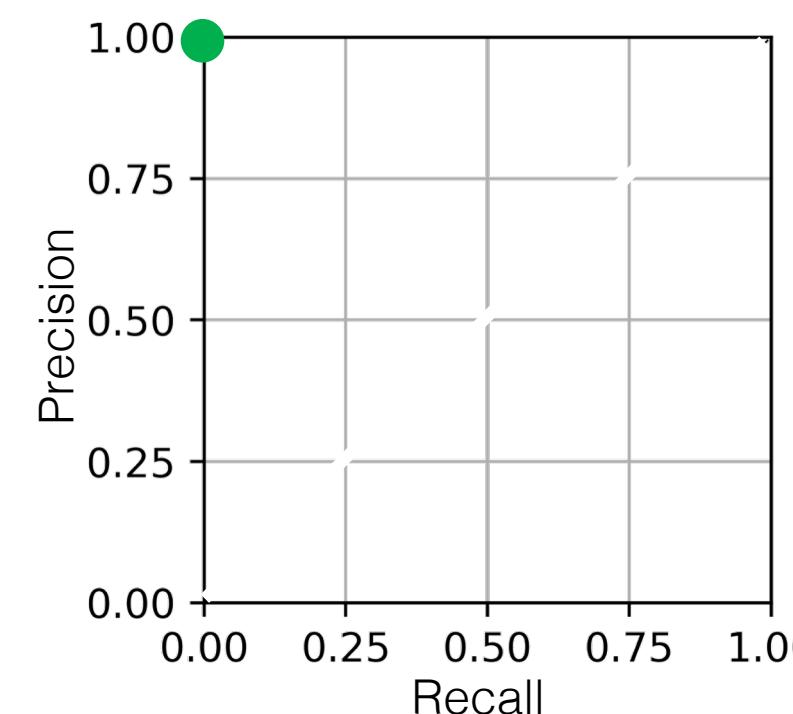
# ROC Curves: how do they compare?



# ROC Curves: where do we operate?



# PR Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



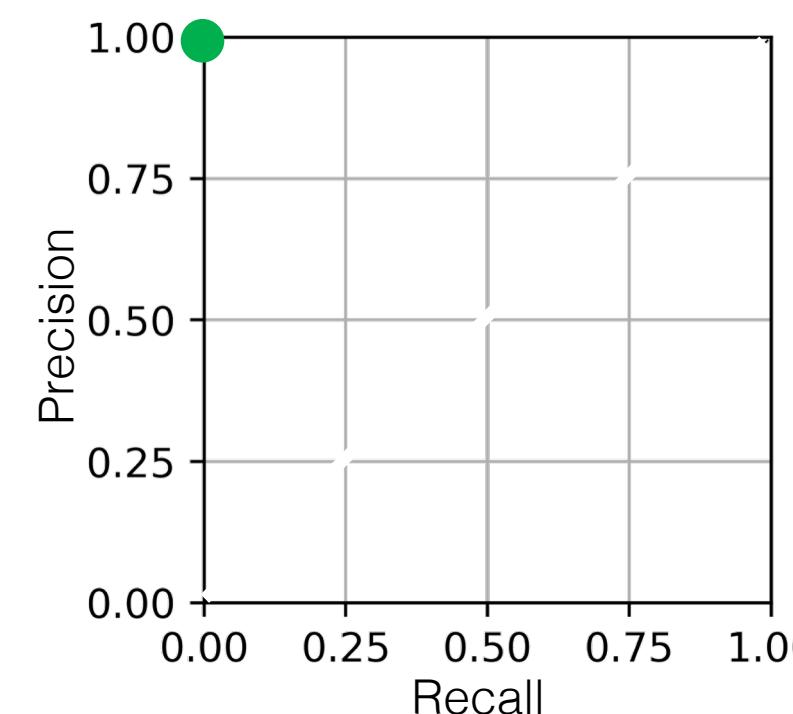
Total Positives = 3

Total Negatives = 2

Threshold	# True Positives	Recall	# Predicted Positive	Precision
-----------	------------------	--------	----------------------	-----------

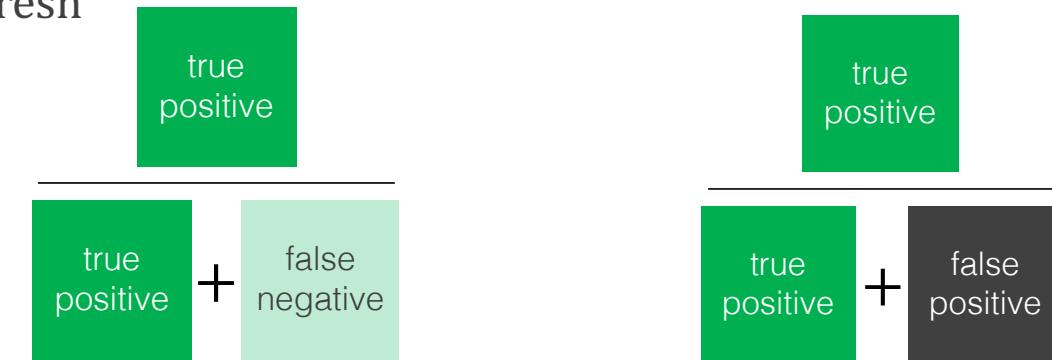
True Class Label (y)	Classifier Confidence
1	1.40
1	0.95
0	0.80
1	0.60
0	-0.10

# PR Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



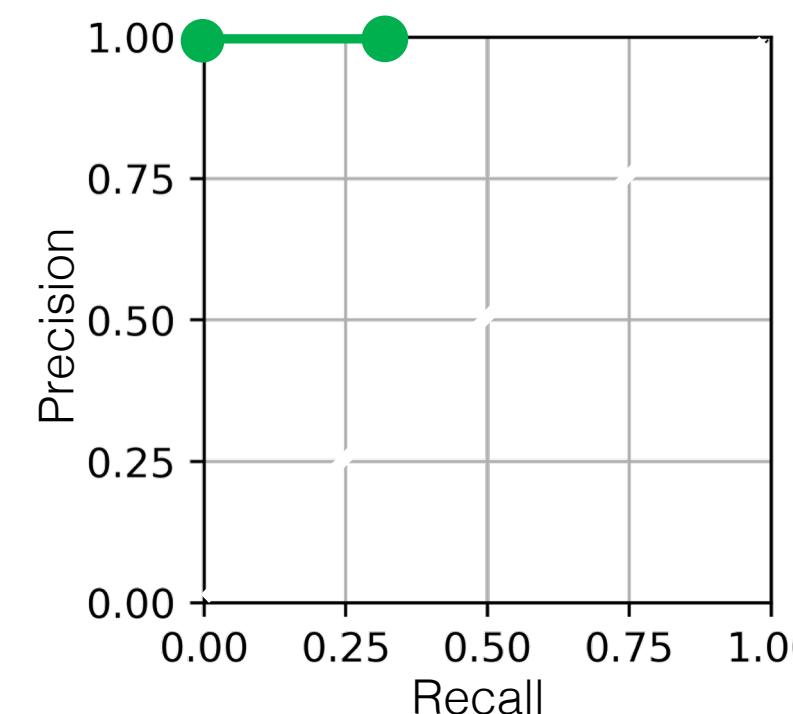
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
0	1	1.40
0	1	0.95
0	0	0.80
0	1	0.60
0	0	-0.10

Threshold	# True Positives	Recall	# Predicted Positive	Precision
$\infty$	0	0	0	undefined

# PR Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



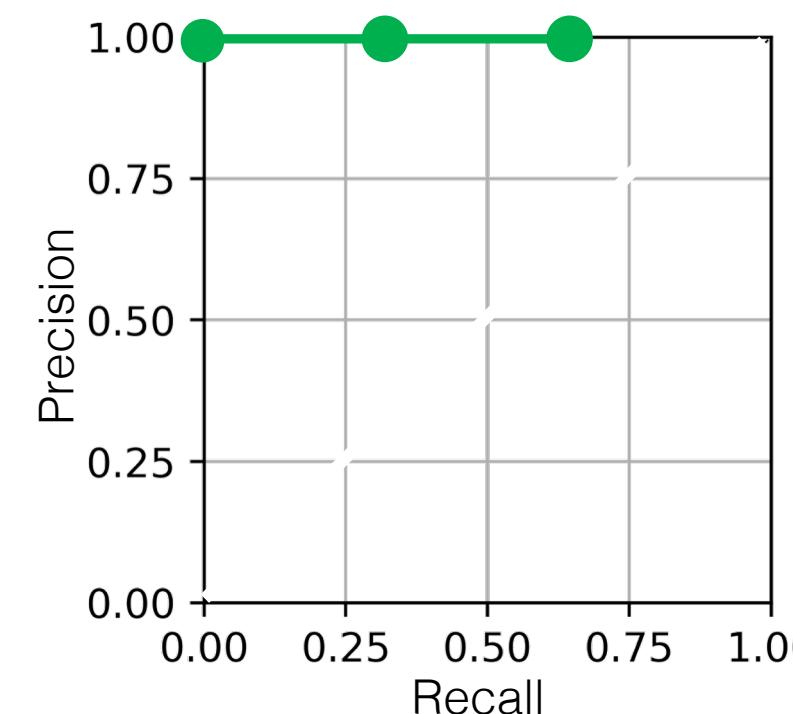
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
0	1	0.95
0	0	0.80
0	1	0.60
0	0	-0.10

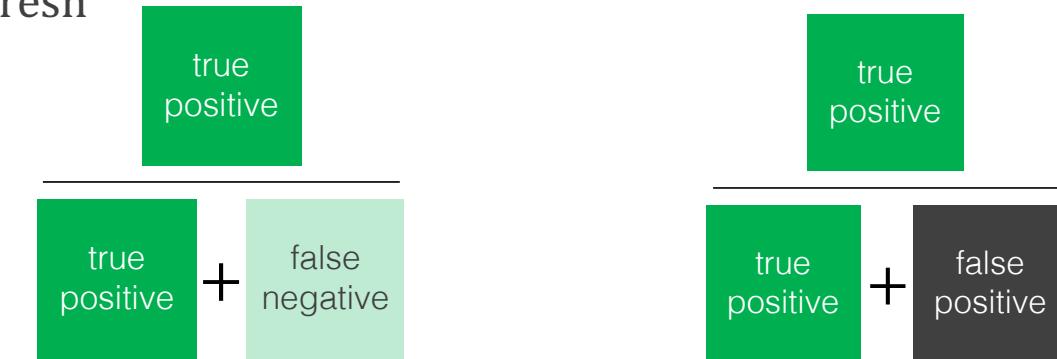
Threshold	# True Positives	Recall	# Predicted Positive	Precision
$\infty$	0	0	0	undefined
1.0	1	0.333	1	1

# PR Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



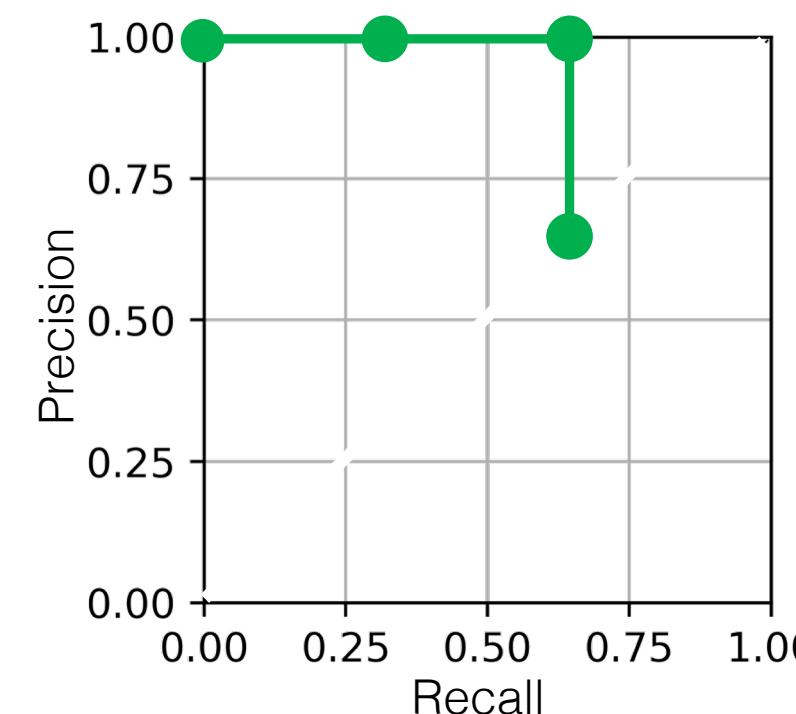
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
0	0	0.80
0	1	0.60
0	0	-0.10

Threshold	# True Positives	Recall	# Predicted Positive	Precision
$\infty$	0	0	0	undefined
1.0	1	0.333	1	1
0.9	2	0.667	2	1

# PR Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



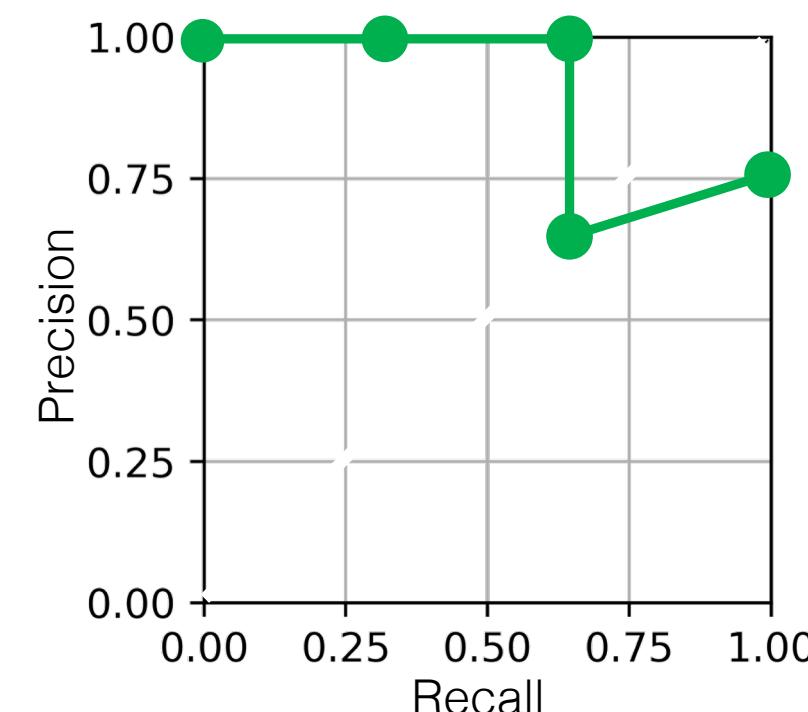
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
1	0	0.80
0	1	0.60
0	0	-0.10

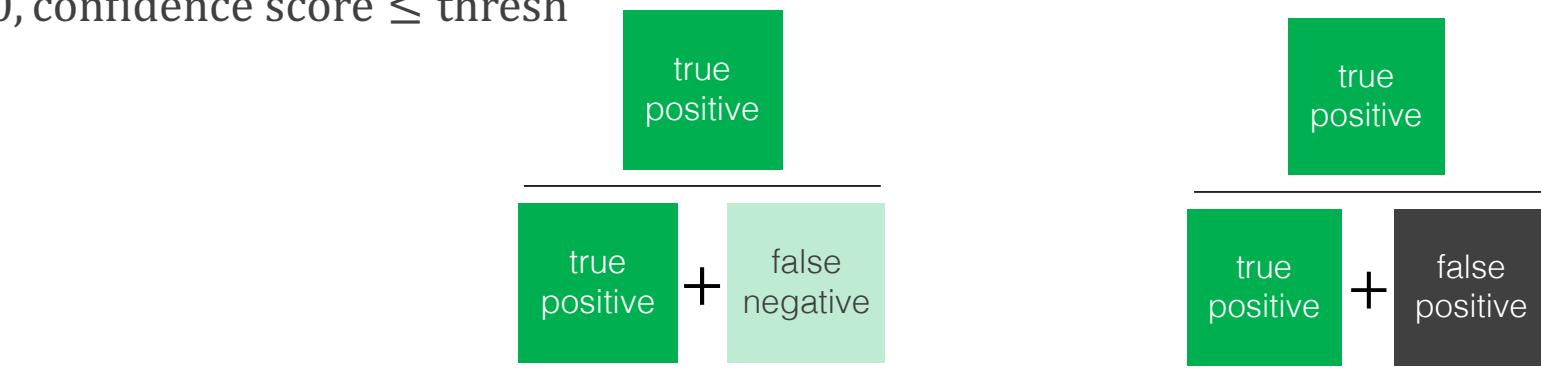
Threshold	# True Positives	Recall	# Predicted Positive	Precision
$\infty$	0	0	0	undefined
1.0	1	0.333	1	1
0.9	2	0.667	2	1
0.7	2	0.667	3	0.667

# PR Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



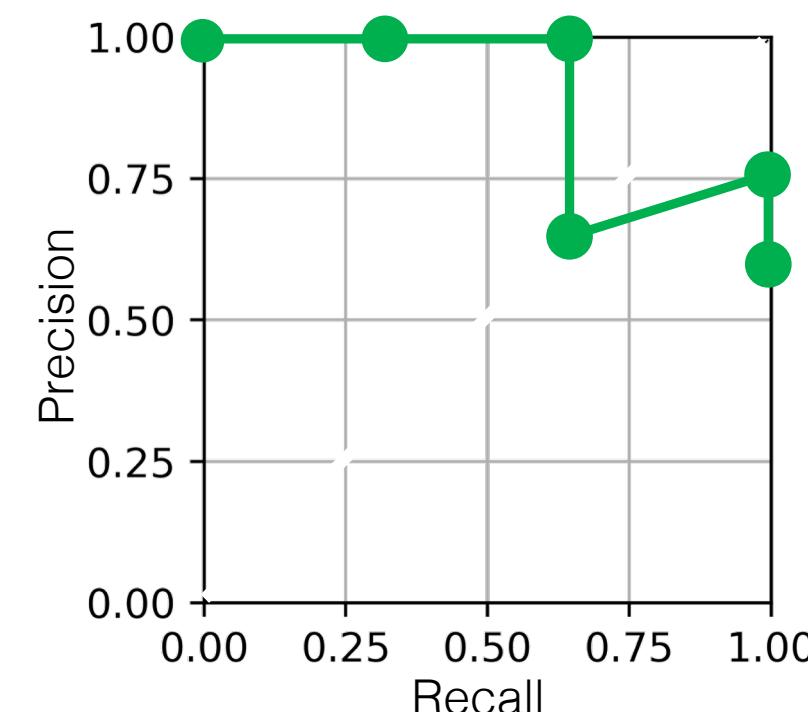
Total Positives = 3

Total Negatives = 2

Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
1	0	0.80
1	1	0.60
0	0	-0.10

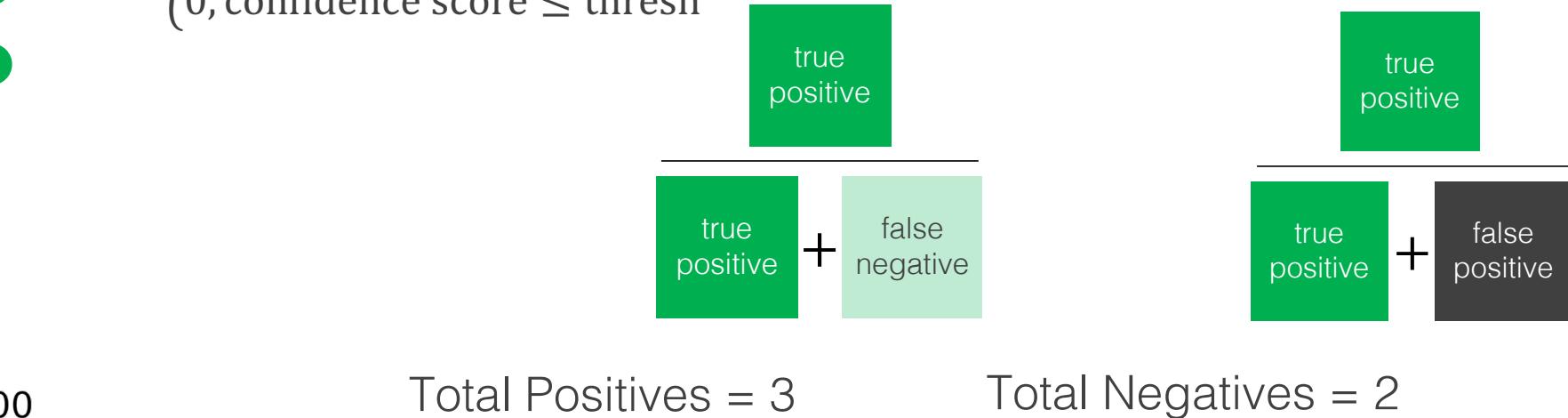
Threshold	# True Positives	Recall	# Predicted Positive	Precision
$\infty$	0	0	0	undefined
1.0	1	0.333	1	1
0.9	2	0.667	2	1
0.7	2	0.667	3	0.667
0.0	3	1	4	0.75

# PR Curves



Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$



Estimate ( $\hat{y}$ )	True Class Label (y)	Classifier Confidence
1	1	1.40
1	1	0.95
1	0	0.80
1	1	0.60
1	0	-0.10

Threshold	# True Positives	Recall	# Predicted Positive	Precision
$\infty$	0	0	0	undefined
1.0	1	0.333	1	1
0.9	2	0.667	2	1
0.7	2	0.667	3	0.667
0.0	3	1	4	0.75
$-\infty$	3	1	5	0.6

# Case study 1

$i$	$y_i$	$\hat{y}_i$
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	0
8	0	1
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0

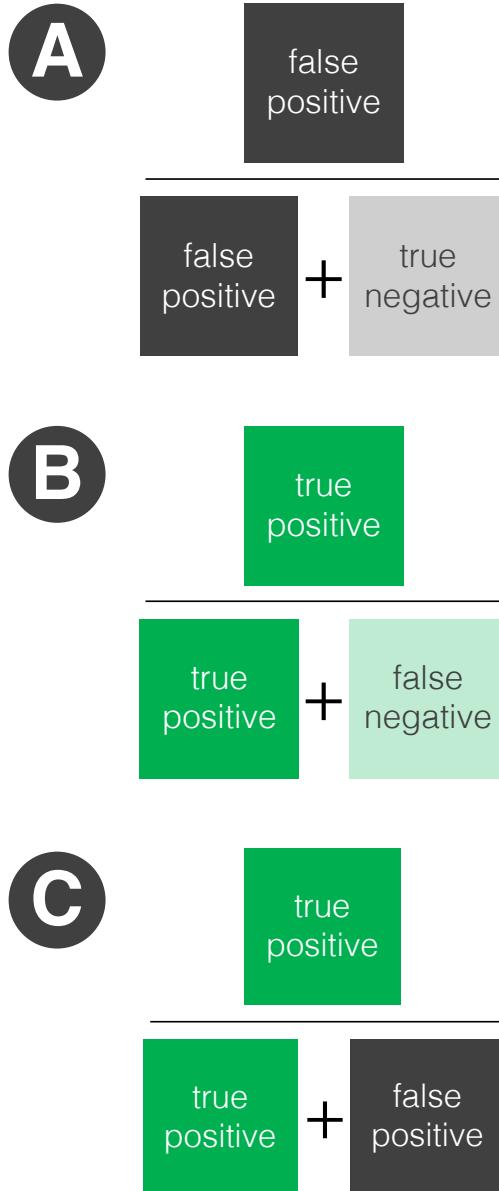
**Overall classification accuracy =  $13/15 = 0.87$**

**ROC Curves** measure the tradeoff between...

- A** False positive rate =  $1/8 = 0.13$
- B** True positive rate (Recall) =  $6/7 = 0.86$

**PR Curves** measure the tradeoff between...

- B** True positive rate (Recall) =  $6/7 = 0.86$
- C** Precision=  $6/7 = 0.86$



# Case study II

$i$	$y_i$	$\hat{y}_i$
1	1	1
2	1	1
3	1	0
4	1	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0

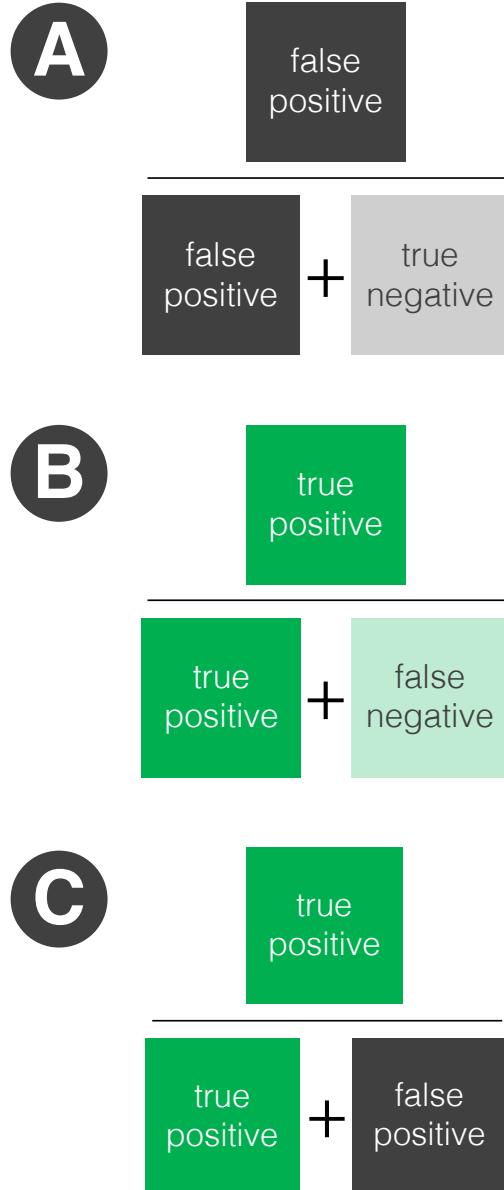
Overall classification accuracy =  $13/15 = 0.87$

ROC Curves measure the tradeoff between...

- A False positive rate =  $0/11 = 0$
- B True positive rate (Recall) =  $2/4 = 0.5$

PR Curves measure the tradeoff between...

- B True positive rate (Recall) =  $2/4 = 0.5$
- C Precision=  $2/2 = 1$



# Case study III

$i$	$y_i$	$\hat{y}_i$
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	0	1
15	0	1

Overall classification accuracy =  $13/15 = 0.87$

ROC Curves measure the tradeoff between...

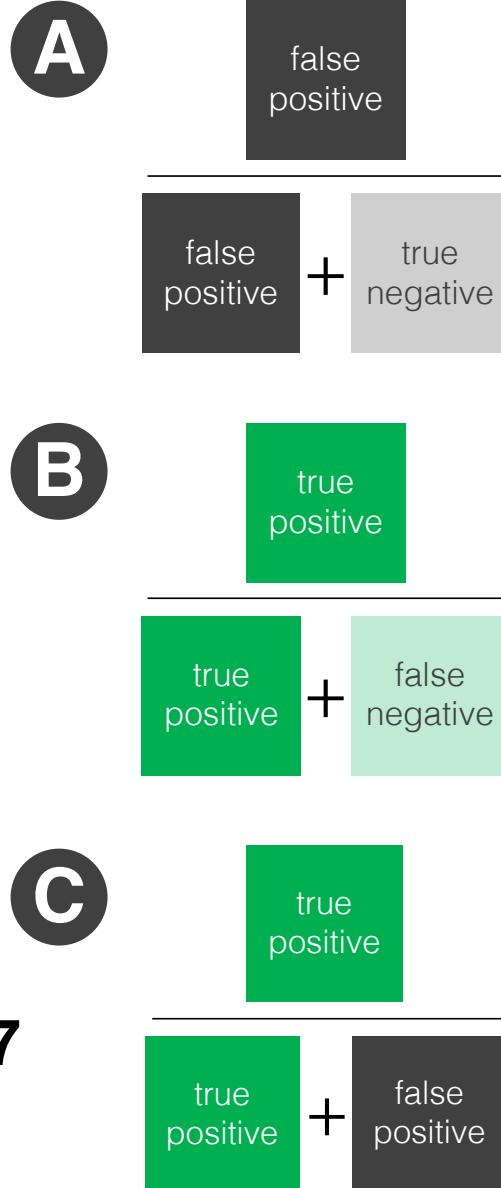
A False positive rate =  $2/2 = 1$

B True positive rate (Recall) =  $13/13 = 1$

PR Curves measure the tradeoff between...

B True positive rate (Recall) =  $13/13 = 1$

C Precision=  $13/15 = 0.87$



# Multiclass Classification: Confusion Matrix

Predicted Class, $\hat{y}$			
Class 1	Class 2	Class 3	
True Class, $y$	Class 1	Class 2	
Class 1	190	8	2
Class 2	1	5	4
Class 3	24	24	25

No. samples  
from class

[200]

[10]

[73]

confusion matrix with number of samples

# F<sub>1</sub>-score

$$F_1 = 2 \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}}$$

Harmonic mean of precision and recall

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

Generally:

$$F_\beta = (1 + \beta^2) \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

$\beta$  controls the relative weight of precision/recall

# Multiclass F<sub>1</sub>

**Micro-average:** Calculate precision and recall metrics globally by counting the total true positives, false negatives, and false positives  
(average for the whole dataset)

**Macro-average:** Use the average precision and recall for each class label  
(average of class-averages)