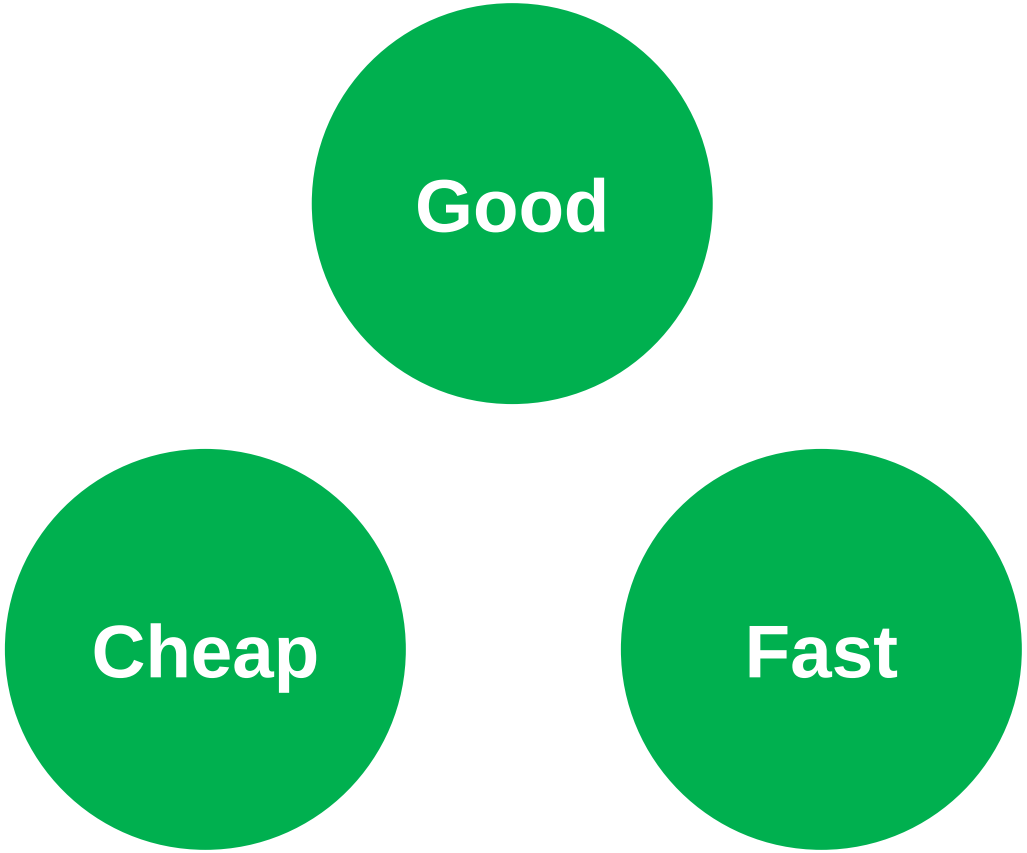
Evaluating Performance I

Lecture 06

Choose 2

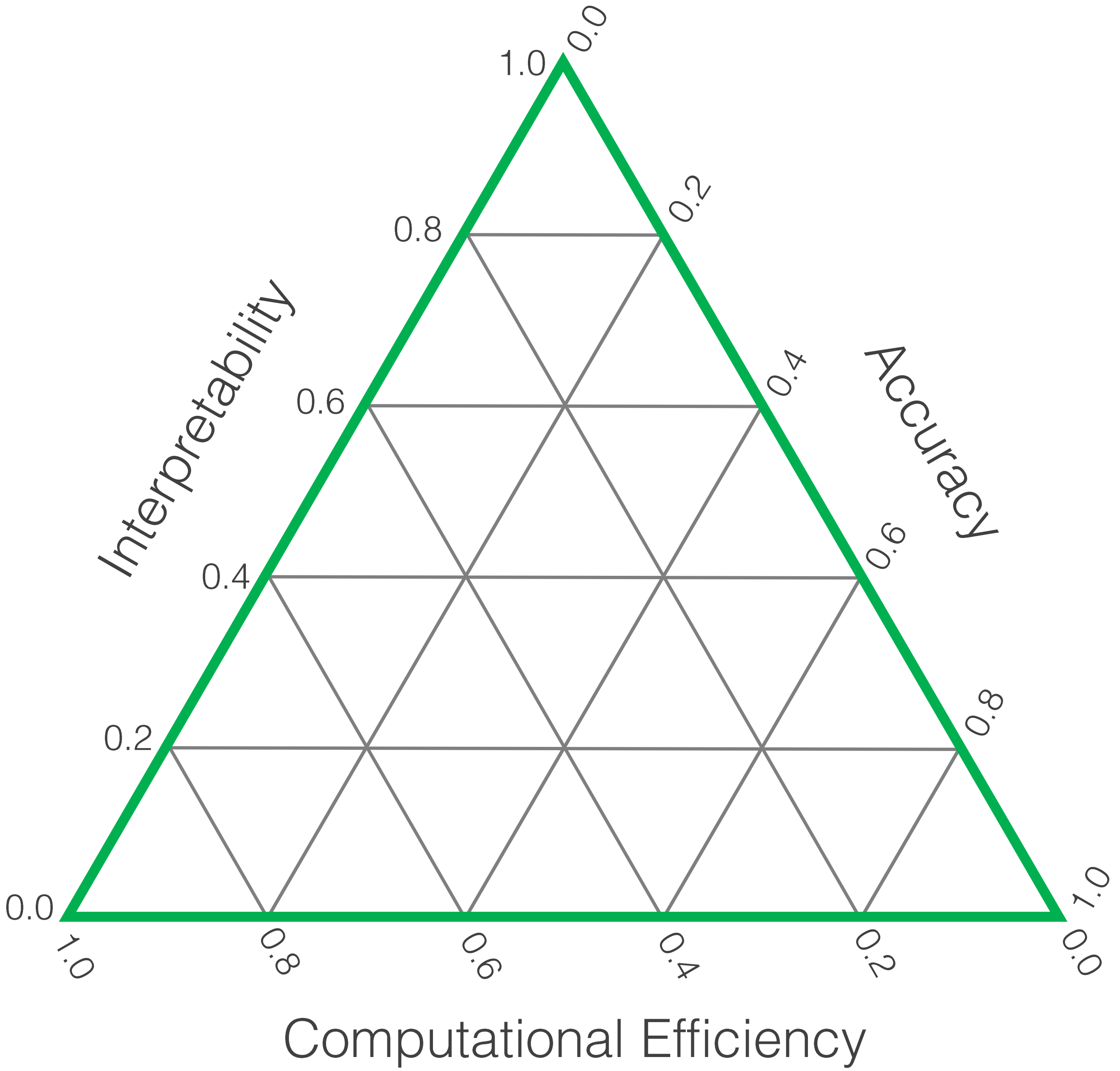


Modeling Tradeoffs

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

- Classification accuracy
- False alarm rate
- Recall
- Precision
- F_1 Score
- Area under the ROC curve (AUC)

Multiclass

Confusion matrices

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

Common Metrics

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^2 Coefficient of determination

R^2 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

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

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

Relative measure of performance

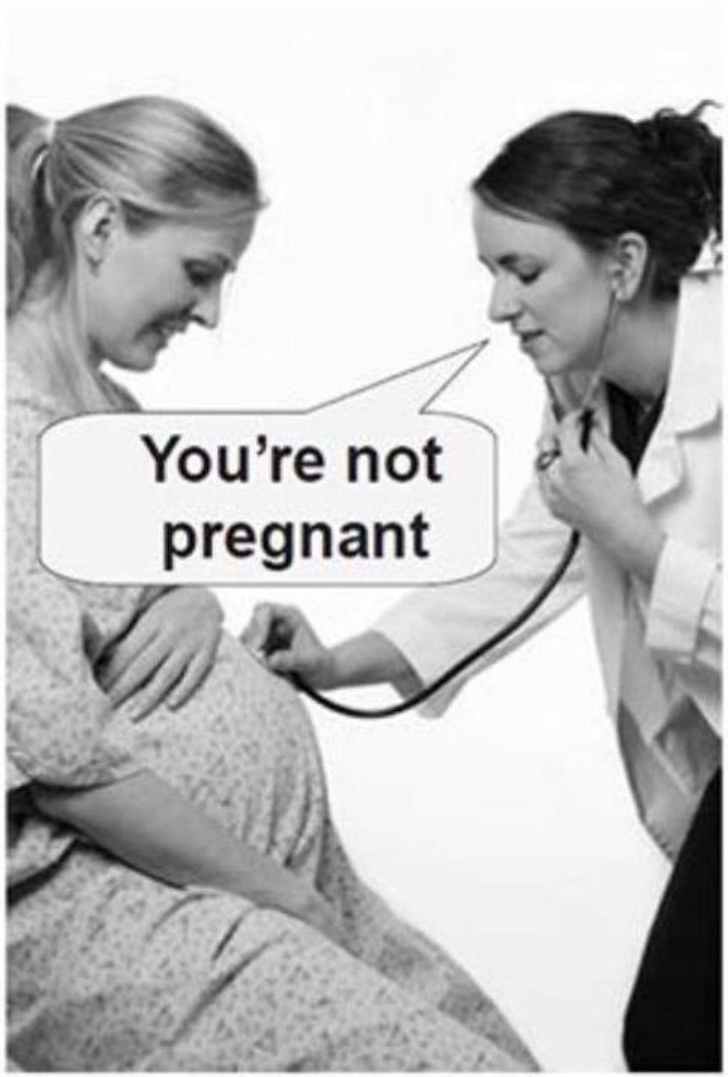
Types of error

False Positive

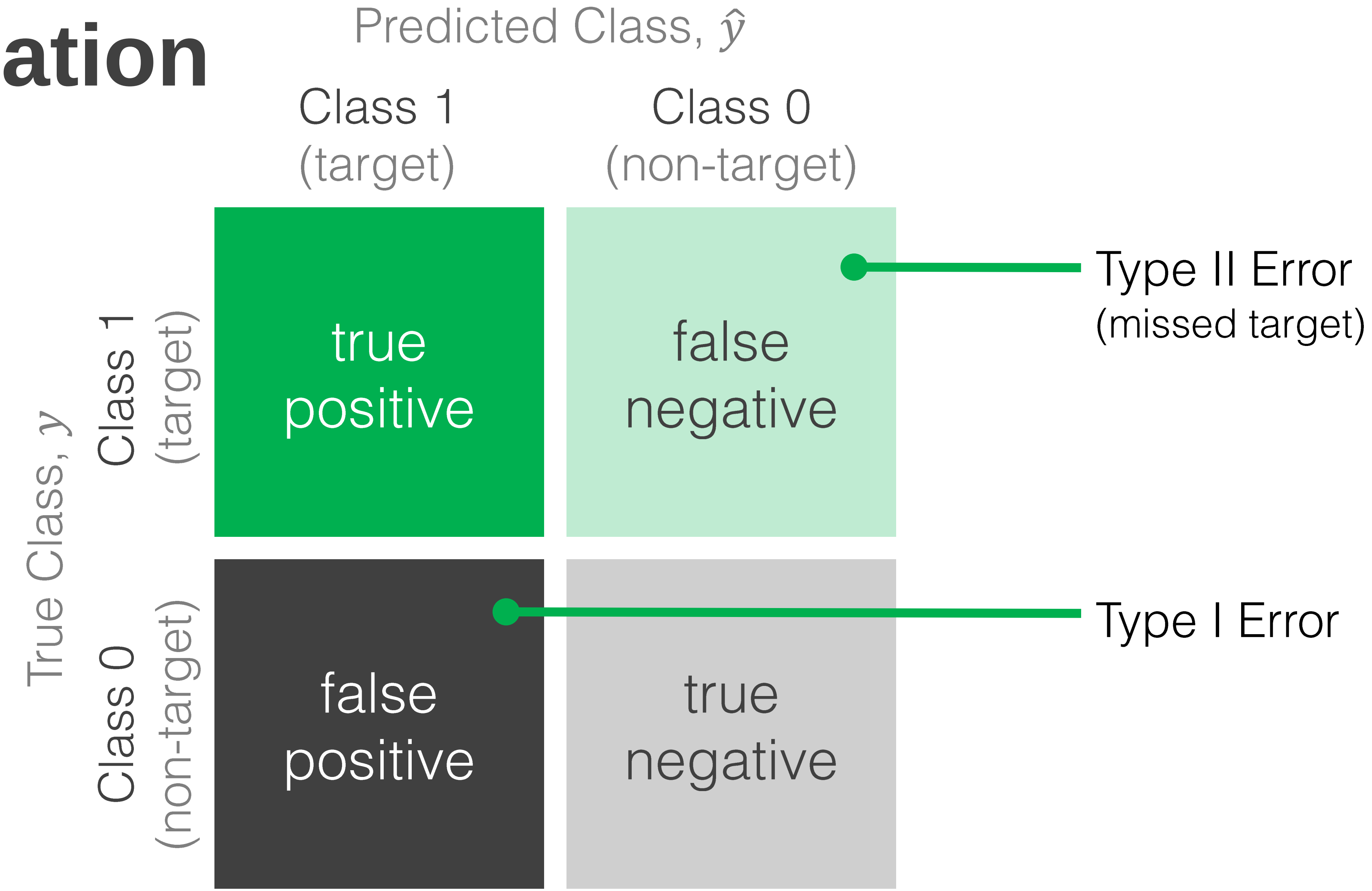
(Type I error)

False Negative

(Type II error)



Binary Classification



Binary Classification

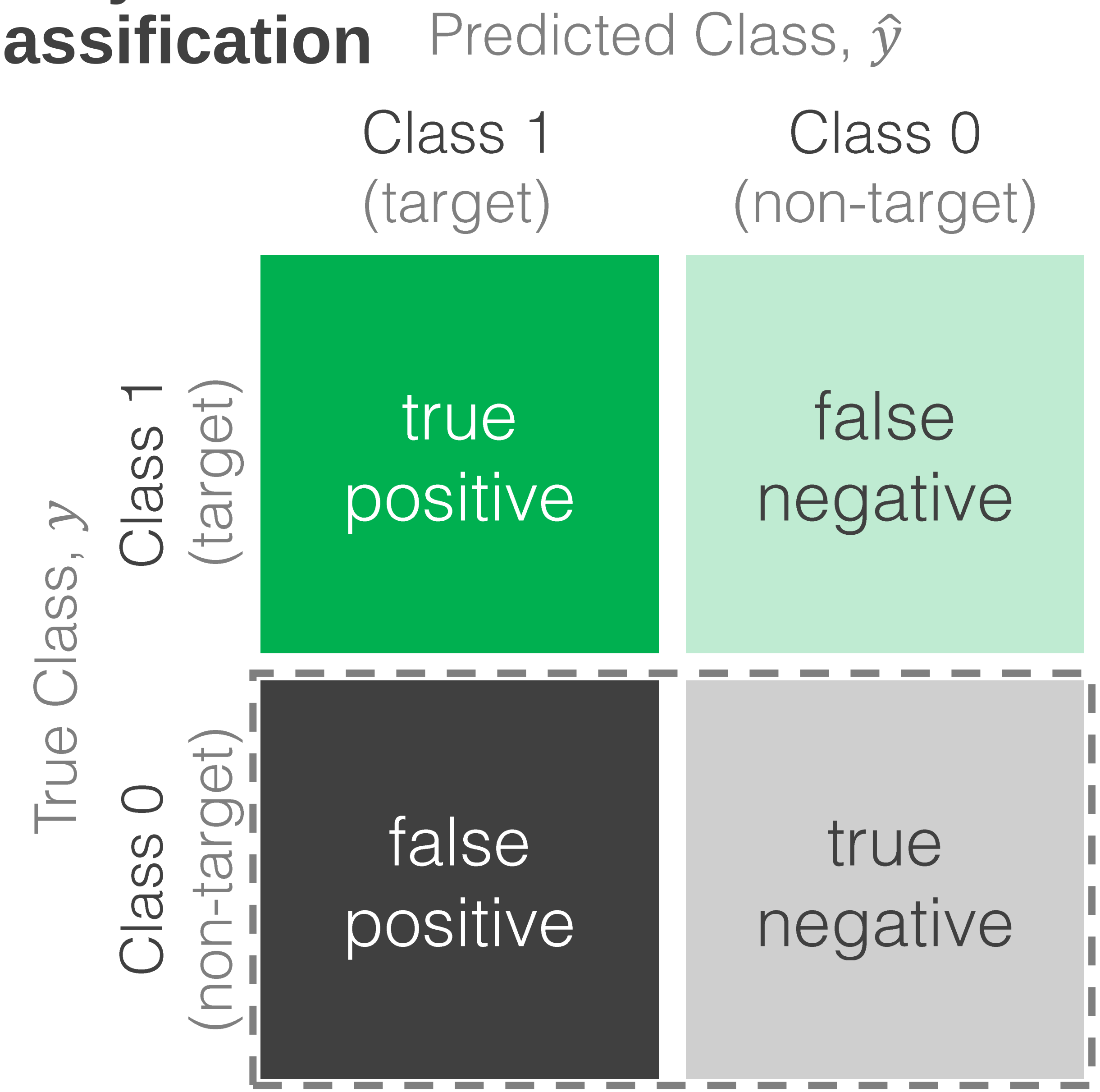
True Class, y \ Predicted Class, $\hat{y}$	Class 1 (target)	Class 0 (non-target)
Class 1 (target)	true positive	false negative
Class 0 (non-target)	false positive	true negative

True positive rate
Probability of detection, p_D
Sensitivity
Recall

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

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

Binary Classification



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

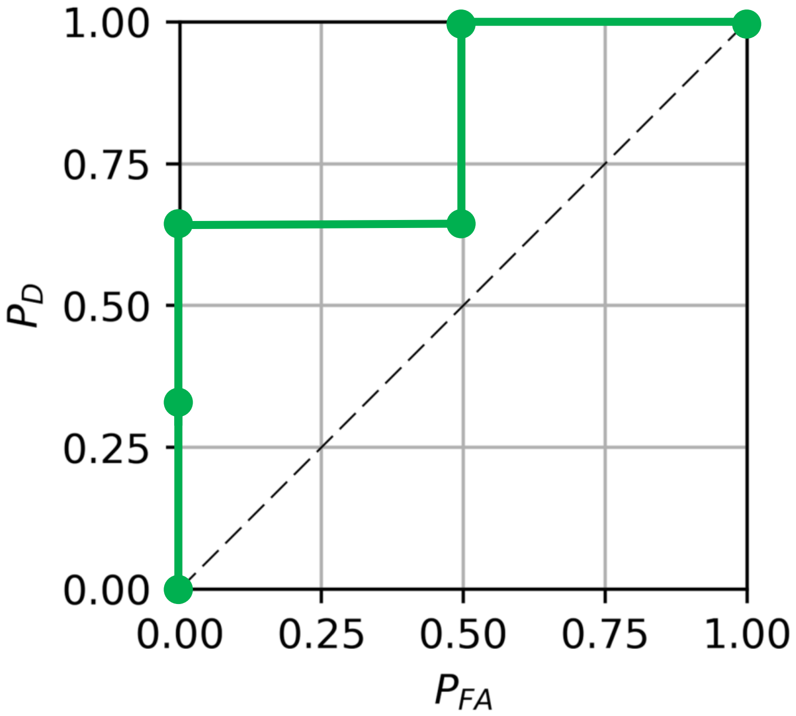
		Predicted Class, $\hat{y}$: Class 1 (target)	Predicted Class, $\hat{y}$: Class 0 (non-target)
True Class, y	Class 1 (target)	true positive	false negative
True Class, y	Class 0 (non-target)	false positive	true negative

Precision

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

How many of the predicted targets are targets?

ROC Curves



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

Classifier decision rule:

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

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

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

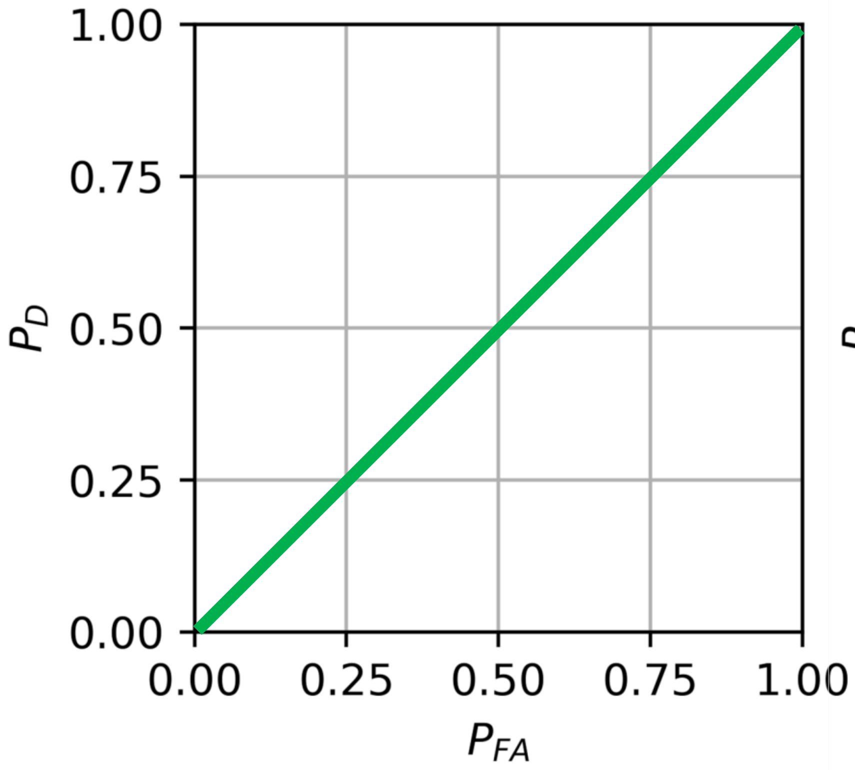
$$\frac{\text{false positive}}{\text{false positive} + \text{true negative}}$$

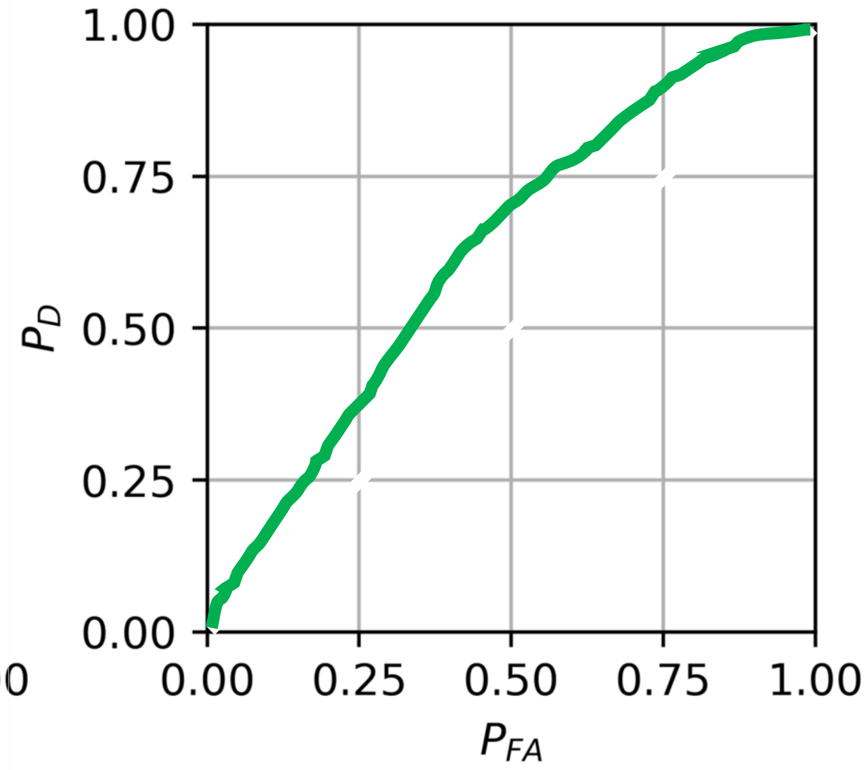
Total Positives = 3

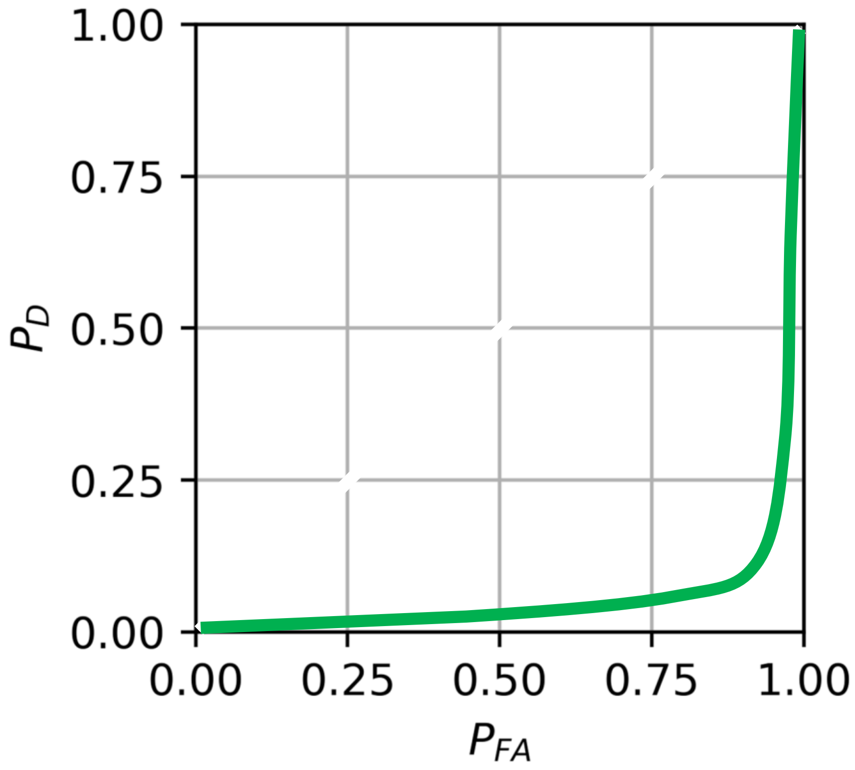
Total Negatives = 2

Threshold	# True Positives	Correct detection rate, P_D	# False Positives	False alarm rate, P_{FA}

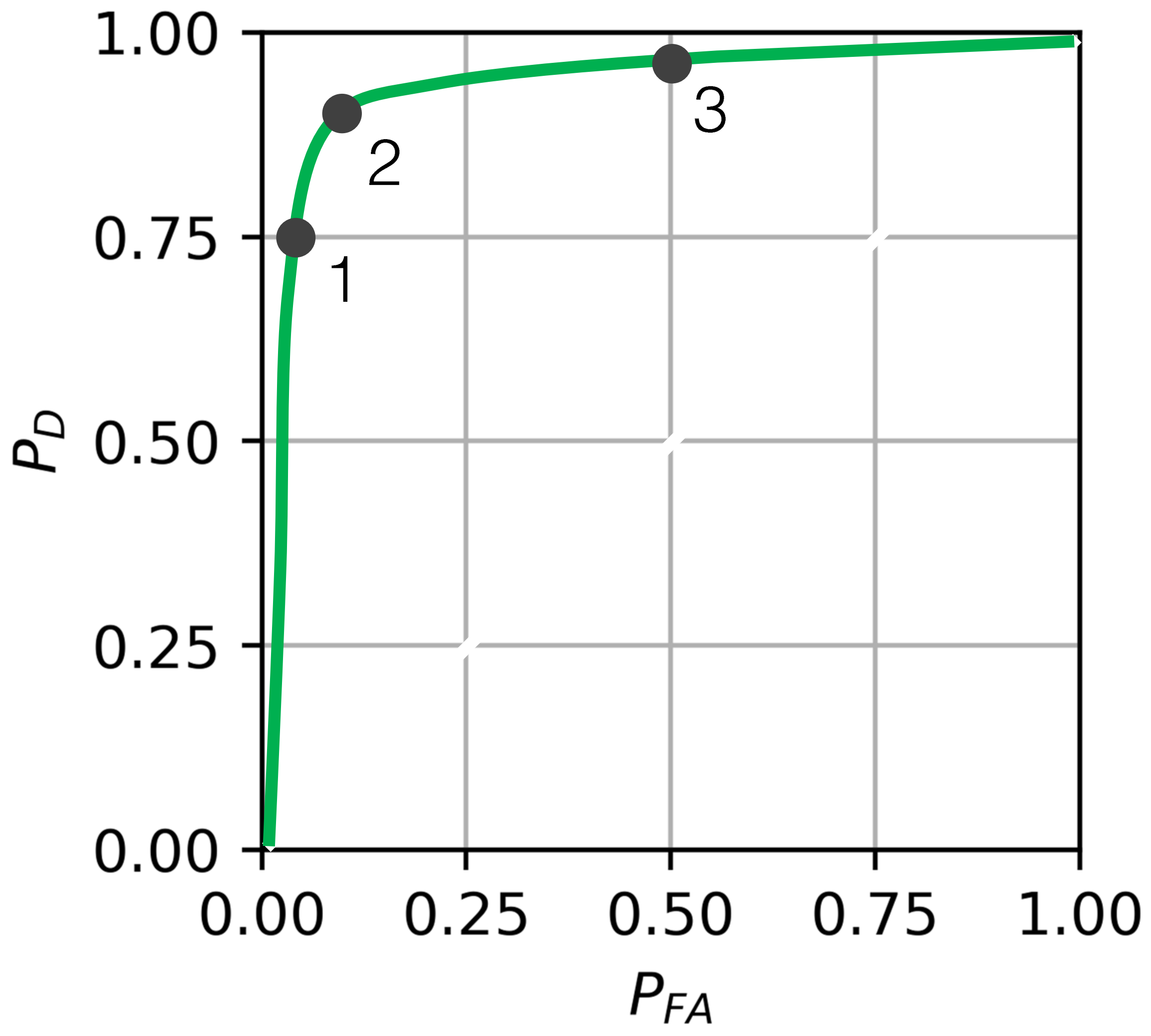
ROC Curves: how do they compare?



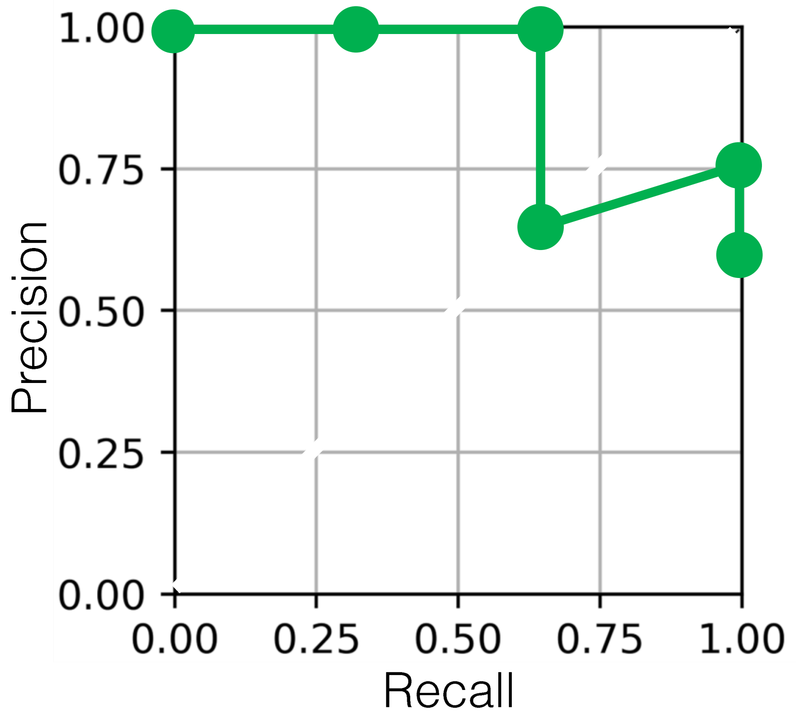




ROC Curves: where do we operate?

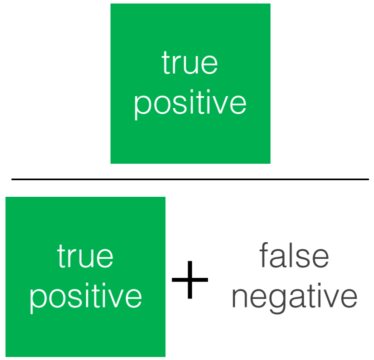


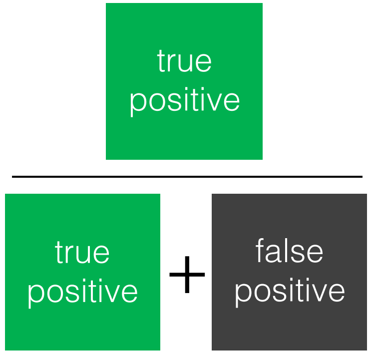
PR Curves



Classifier decision rule:

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





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

Total Positives = 3 Total Negatives = 2

Threshold	# True Positives	Recall, P_D	# Predicted Positive	Precision

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**

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

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

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

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**

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

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

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

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**

A $\dfrac{\text{false positive}}{\text{false positive} + \text{true negative}}$

B $\dfrac{\text{true positive}}{\text{true positive} + \text{false negative}}$

C $\dfrac{\text{true positive}}{\text{true positive} + \text{false positive}}$

Multiclass Classification: Confusion Matrix

Predicted Class, $\hat{y}$

True Class, y	Class 1	Class 2	Class 3	No. samples from class
Class 1	190	8	2	[200]
Class 2	1	5	4	[10]
Class 3	24	24	25	[73]

confusion matrix with number of samples

Predicted Class, $\hat{y}$

True Class, y	Class 1	Class 2	Class 3	
Class 1	0.95	0.04	0.01	[200]
Class 2	0.10	0.50	0.40	[10]
Class 3	0.33	0.33	0.34	[73]

confusion matrix with probabilities

F_1-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:

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

β controls the relative weight of precision/recall

Multiclass F_1

Micro-average: Calculate metrics globally by counting the total true positives, false negatives, and false positives

Macro-average: Use the average precision and recall for each class label