

# DATA SCIENCE

## MODEL EVALUATION METRICS

### *Regression:*

- *Root Mean Squared Error*

### *Classification:*

- *Confusion Matrix*
- *ROC Curve (and AUC)*

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- *Used for regression problems*
- *Square root of the mean of the squared errors*
- *Easily interpretable (in the “y” units)*
- *“Punishes” larger errors*

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

*Example:*

*$y\_true = [100, 50, 30]$*

*$y\_preds = [90, 50, 50]$*

*$RMSE = np.sqrt((10**2 + 0**2 + 20**2) / 3) = 12.88$*

*Confusion Matrix: table to describe the performance of a classifier*

n=165	Predicted: NO	Predicted: YES
	Actual: NO	Actual: YES
	50	10
	5	100

*Example: Test for presence of disease*

*NO = negative test = False = 0*

*YES = positive test = True = 1*

- *How many classes are there?*
- *How many patients?*
- *How many times is disease predicted?*
- *How many patients actually have the disease?*

# CONFUSION MATRIX

6

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

## *Basic Terminology:*

- *True Positives (TP)*
- *True Negatives (TN)*
- *False Positives (FP)*
- *False Negatives (FN)*

## *Accuracy:*

- *Overall, how often is it **correct**?*
- $(TP + TN) / total = 150 / 165 = 0.91$

## *Misclassification Rate (Error Rate):*

- *Overall, how often is it **wrong**?*
- $(FP + FN) / total = 15 / 165 = 0.09$

# CONFUSION MATRIX

7

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

## *False Positive Rate:*

- When actual value is **negative**, how often is prediction **wrong**?
- $FP / \text{actual no} = 10/60 = 0.17$

## *Sensitivity:*

- When actual value is **positive**, how often is prediction **correct**?
- $TP / \text{actual yes} = 100/105 = 0.95$
- “True Positive Rate” or “Recall”

## *Specificity:*

- When actual value is **negative**, how often is prediction **correct**?
- $TN / \text{actual no} = 50/60 = 0.83$

Email Number	Score	True Label
5	0.99	Spam
8	0.82	Spam
2	0.60	Spam
1	0.60	Ham
7	0.48	Spam
3	0.22	Ham
4	0.10	Ham
6	0.02	Ham

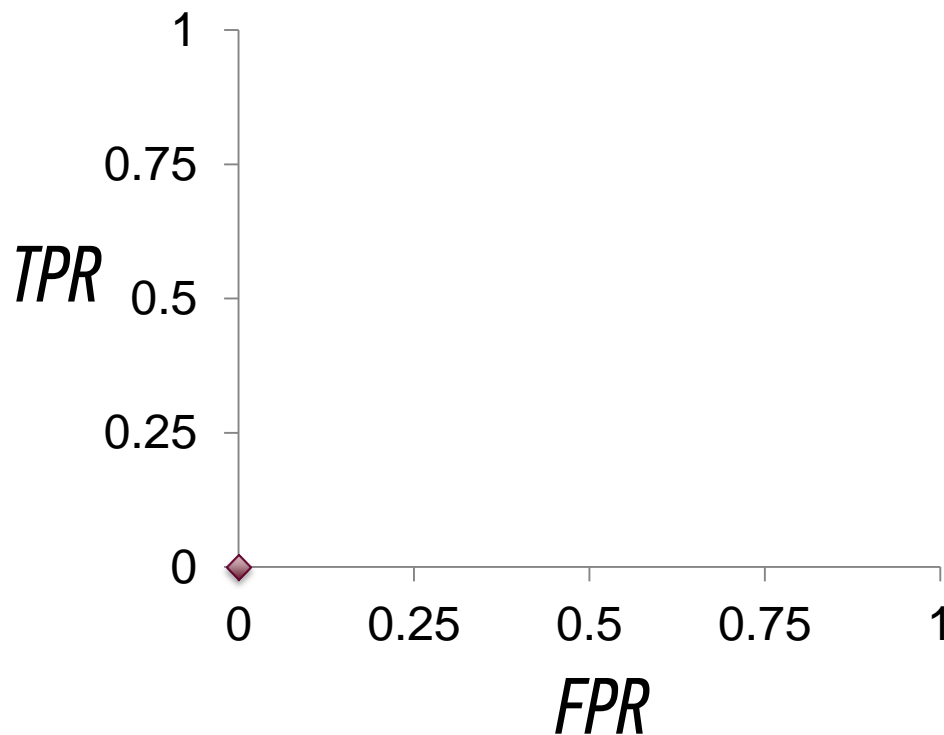
*Every email is assigned a “spamminess” score by our classification algorithm. To actually make our predictions, we choose a numeric cutoff for classifying as spam.*

*An ROC Curve will help us to visualize how well our classifier is doing without having to choose a cutoff!*



Email Number	Score	True Label
5	0.99	Spam
8	0.82	Spam
2	0.60	Spam
1	0.60	Ham
7	0.48	Spam
3	0.22	Ham
4	0.10	Ham
6	0.02	Ham

*ROC Curve*



Email Number	Score	True Label
5	0.99	Spam
8	0.82	Spam
2	0.60	Spam
1	0.60	Ham
7	0.48	Spam
3	0.22	Ham
4	0.10	Ham
6	0.02	Ham

TPR: When actual value is **spam**, how often is prediction **correct**?

FPR: When actual value is **ham**, how often is prediction **wrong**?

Cutoff	TPR (y)	FPR (x)	Cutoff	TPR (y)	FPR (x)
0			0.50		
0.05			0.65		
0.15			0.85		
0.25			1		

Email Number	Score	True Label
5	0.99	Spam
8	0.82	Spam
2	0.60	Spam
1	0.60	Ham
7	0.48	Spam
3	0.22	Ham
4	0.10	Ham
6	0.02	Ham

TPR: When actual value is **spam**, how often is prediction **correct**?

FPR: When actual value is **ham**, how often is prediction **wrong**?

Cutoff	TPR (y)	FPR (x)	Cutoff	TPR (y)	FPR (x)
<b>0</b>	1	1	<b>0.50</b>	0.75	0.25
<b>0.05</b>	1	0.75	<b>0.65</b>	0.5	0
<b>0.15</b>	1	0.5	<b>0.85</b>	0.25	0
<b>0.25</b>	1	0.25	<b>1</b>	0	0

Email Number	Score	True Label
5	0.99	Spam
8	0.98	Spam
2	0.97	Spam
1	0.97	Ham
7	0.96	Spam
3	0.95	Ham
4	0.94	Ham
6	0.93	Ham

*Q: Would the ROC Curve (and AUC) change if the **scores** changed but the **ordering** remained the same?*

*A: Not at all! The ROC Curve is only sensitive to **rank ordering** and does not require **calibrated scores**.*

---

# DATA SCIENCE

---