



# Machine Learning

## Week 5: Classification Performance Metrics

# Outline

---

*“Numbers have an important story to tell. They rely on you to give them a voice.”—Stephen Few*

- Confusion matrix
- Accuracy
- Precision
- Recall
- F-Score
- Area under ROC

# Data Description

---

- Breast cancer data set: Medical data from 681 women (instances) who has potentially cancerous tumors (12 attributes)

## Class attribute (1)

0: Tumor is malignant (238)

1: Tumor is benign (443)

## Other attributes (2)

PID: Patient ID

Date: Diagnosis Date

## Predictor attributes (9)

Adhes - marginal adhesion

BNucl - bare nuclei

Chrom - bland chromatin

Epith - epithelial cell size

Mitos – mitoses

NNucl - normal nucleoli

Thick - clump thickness

UShap - cell shape uniformity

USize - cell size uniformity

\* A predictor is assigned the value 1 if it is normal and the value 10 if it is most abnormal

# Sample Dataset

---

PID	Date	Adhes	BNucl	Chrom	Epith	Mitos	NNucl	Thick	UShape	USize	Class
1	01/03/2007	1	1	3	2	1	1	5	1	1	1
2	12/12/2005	5	10	3	7	1	2	5	4	4	1
3	14/08/2016	1	2	3	2	1	1	3	1	1	1
4	02/02/2001	1	4	3	3	1	7	6	8	8	1
5	14/11/2014	3	1	3	2	1	1	4	1	1	1
6	22/09/2011	8	10	9	7	1	7	8	10	10	0
7	18/05/2015	7	8	8	9	2	8	8	6	6	0
8	27/04/2011	5	9	9	10	2	6	7	9	9	0
9	19/02/2003	8	6	8	3	1	3	5	10	10	0
10	25/07/2011	10	5	6	6	4	4	10	7	7	0

# Prepare Data

---

Remove  
attributes. Not  
useful  
predictors.

PID	Date	Adhes	BNucl	Chrom	Epith	Mitos	NNucl	Thick	UShape	USize	Class
1	01/03/2007	1	1	3	2	1	1	5	1	1	1
2	12/12/2005	5	10	3	7	1	2	5	4	4	1
3	14/08/2016	1	2	3	2	1	1	3	1	1	1
4	02/02/2001	1	4	3	3	1	7	6	8	8	1
5	14/11/2014	3	1	3	2	1	1	4	1	1	1
6	22/09/2011	8	10	9	7	1	7	8	10	10	0
7	18/05/2015	7	8	8	9	2	8	8	6	6	0
8	27/04/2011	5	9	9	10	2	6	7	9	9	0
9	19/02/2003	8	6	8	3	1	3	5	10	10	0
10	25/07/2011	10	5	6	6	4	4	10	7	7	0

Class attribute

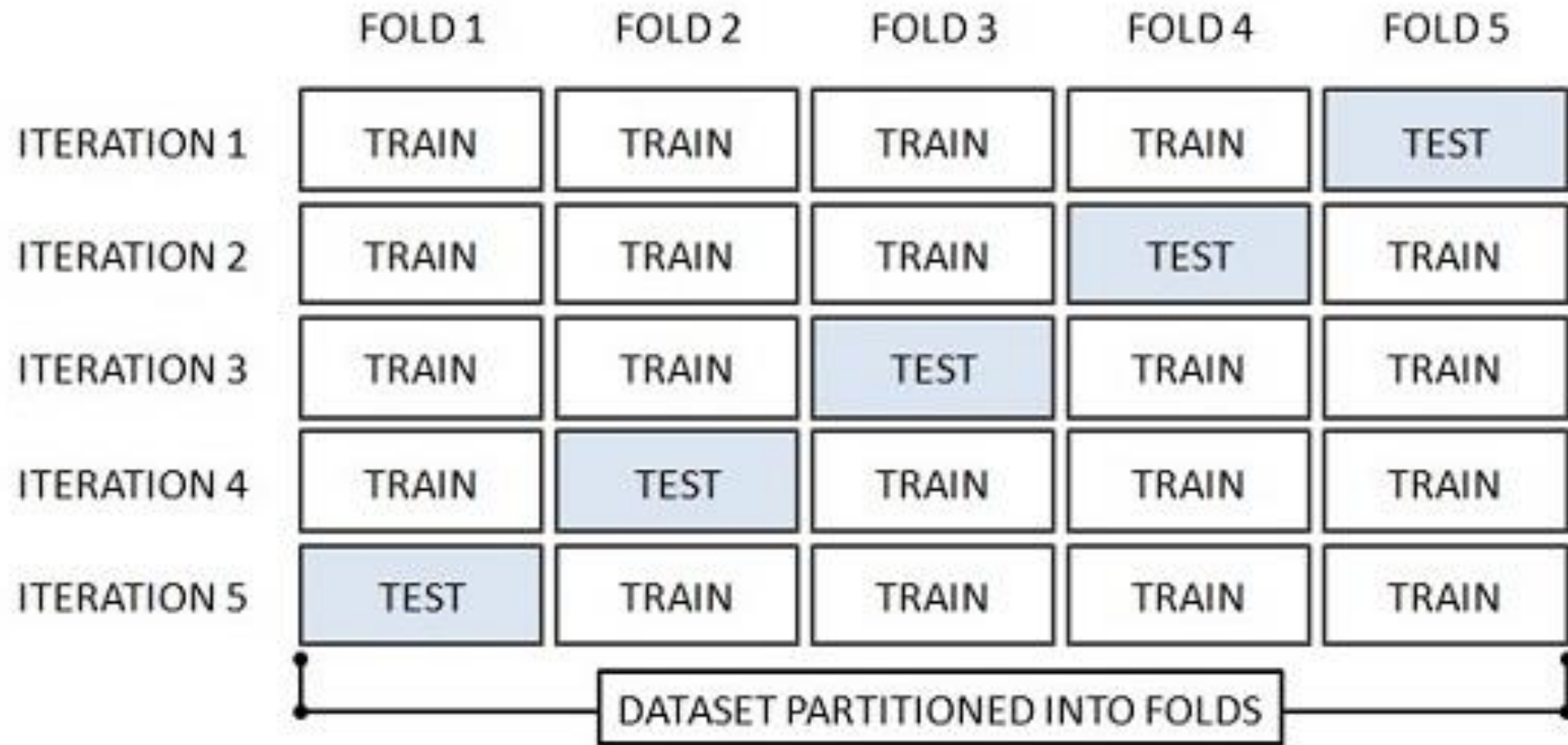
Predictor attributes

# Test Options

---

- Set up train and test sets
  - **Percentage split:** Splits the data and separates  $x\%$  of the data for training and the rest for testing
  - **Supplied test set:** Prepare own external file as training set
  - **K-fold cross validation:** Data set is divided into  $k$  subsets. Each time, one of the  $k$  subsets is used as the test set and the other  $k-1$  subsets are put together to form a training set. Then the average error across all  $k$  trials is computed.

# Cross Validation



5-fold  
cross  
validation

# Confusion Matrix

---

N = 10	Predicted Class			Total
Actual Class		0	1	
	0: Malignant	<b>3 (TP)</b>	<b>2 (FN)</b>	5
	1: Benign	<b>1 (FP)</b>	<b>4 (TN)</b>	5
Total		4	6	10

**True Positives (TP):** Actual class of the data point was TRUE and the predicted is also TRUE (positive class)

*Ex: The case where a tumor is malignant and the model classifying the tumor as malignant*

**True Negatives (TN):** Actual class of the data point was FALSE and the predicted is also FALSE (negative class)

*Ex: The case where the tumor is benign and the model classifying the tumor as benign*

**False Positives (FP):** Actual class of the data point was FALSE and the predicted is TRUE.

*Ex: A tumor being benign and the model classifying the tumor as malignant*

**False Negatives (FN):** Actual class of the data point was TRUE and the predicted is FALSE.

*Ex: Tumor is malignant and the model classifying the tumor as benign*



# Accuracy

---

N = 10	Predicted Class			Total
Actual Class		0	1	
	0: Malignant	<b>3 (TP)</b>	<b>2 (FN)</b>	5
	1: Benign	<b>1 (FP)</b>	<b>4 (TN)</b>	5
Total		4	6	10

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

# Accuracy

---

Good measure when the  
classes in the data are  
nearly balanced.  
Malignant = 5  
Benign = 5

N = 10	Predicted Class			Total
Actual Class		0	1	
	0: Malignant	<b>3 (TP)</b>	<b>2 (FN)</b>	5
	1: Benign	<b>1 (FP)</b>	<b>4 (TN)</b>	5
Total		4	6	10

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\begin{aligned}\text{Accuracy} &= (3 + 4) / (3 + 4 + 1 + 2) \\ &= 7 / 10 \\ &= 0.7\end{aligned}$$

# Accuracy

---

Accuracy is 80% even though the classifier assigned all 10 instances as BENIGN (1)  
Malignant = 2  
Benign = 8

\* NEVER be used as a measure when classes in the data are a majority of one class

N = 10	Predicted Class			Total
		0	1	
Actual Class	0: Malignant	<b>0 (TP)</b>	<b>2 (FN)</b>	2
	1: Benign	<b>0 (FP)</b>	<b>8 (TN)</b>	8
Total		0	10	10

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\begin{aligned}\text{Accuracy} &= (0 + 8) / (0 + 8 + 0 + 2) \\ &= 8 / 10 \\ &= 0.8\end{aligned}$$

# Precision

---

- Measures how good is the model at whatever it predicted
  - Example: Proportion of tumors predicted as malignant, which are actually malignant

$$\text{Precision} = \frac{TP}{TP + FP}$$

N = 10	Predicted Class			Total
		0	1	
Actual Class	0: Malignant	<b>0 (TP)</b>	<b>2 (FN)</b>	2
	1: Benign	<b>1 (FP)</b>	<b>7 (TN)</b>	8
Total		1	9	10

# Precision

---

- Measures how good is the model at whatever it predicted
  - Example: Proportion of tumors predicted as malignant, which are actually malignant

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision

= 0 / (0 + 1)

= 0 / 1

= 0

N = 10	Predicted Class			Total
Actual Class		0	1	
	0: Malignant	0 (TP)	2 (FN)	2
	1: Benign	1 (FP)	7 (TN)	8
Total		1	9	10

# Recall

---

- Measures indicates how good is the model at picking the correct items
- Example: Proportion of actual malignant tumors being predicted by the algorithm as being malignant

$$\text{Recall} = \frac{TP}{TP + FN}$$

N = 10	Predicted Class			Total
		0	1	
Actual Class	0: Malignant	<b>0 (TP)</b>	<b>2 (FN)</b>	2
	1: Benign	<b>1 (FP)</b>	<b>7 (TN)</b>	8
Total		1	9	10

# Recall

---

- Measures indicates how good is the model at picking the correct items
- Example: Proportion of actual malignant tumors being predicted by the algorithm as being malignant

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\begin{aligned}\text{Recall} &= 0 / (0 + 2) \\ &= 0 / 2 \\ &= 0\end{aligned}$$

N = 10	Predicted Class			Total
		0	1	
Actual Class	0: Malignant	0 (TP)	2 (FN)	2
	1: Benign	1 (FP)	7 (TN)	8
Total		1	9	10

# F-Measure (F1)

---

- Harmonic mean of precision and recall
- A single score that represents both precision and recall

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

Say precision = 0.4, recall = 0.7

$$F1 = (2 * 0.4 * 0.7) / (0.4 + 0.7)$$

$$= 0.56 / 1.1$$

$$= 0.51$$

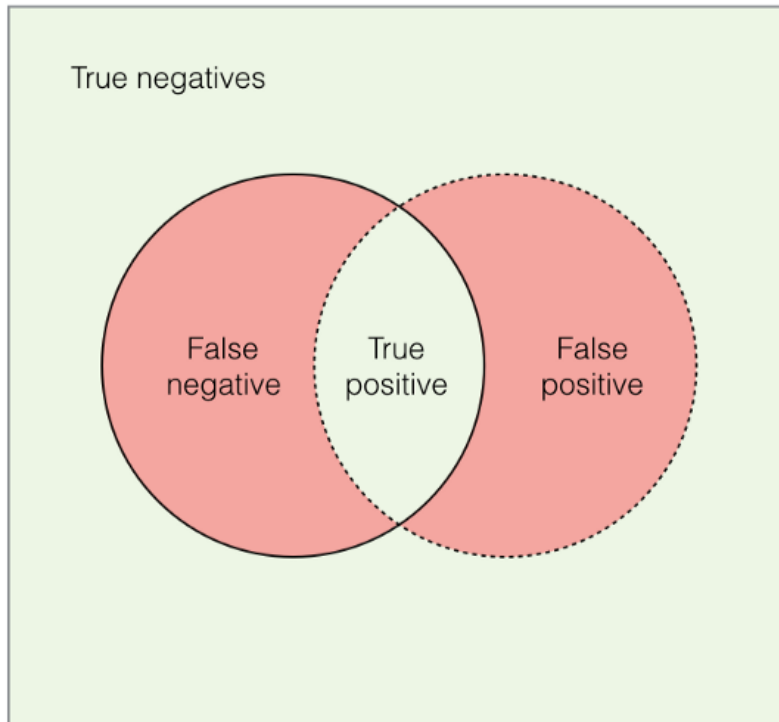


# Precision and Recall

---

**Low Recall, Low Precision**

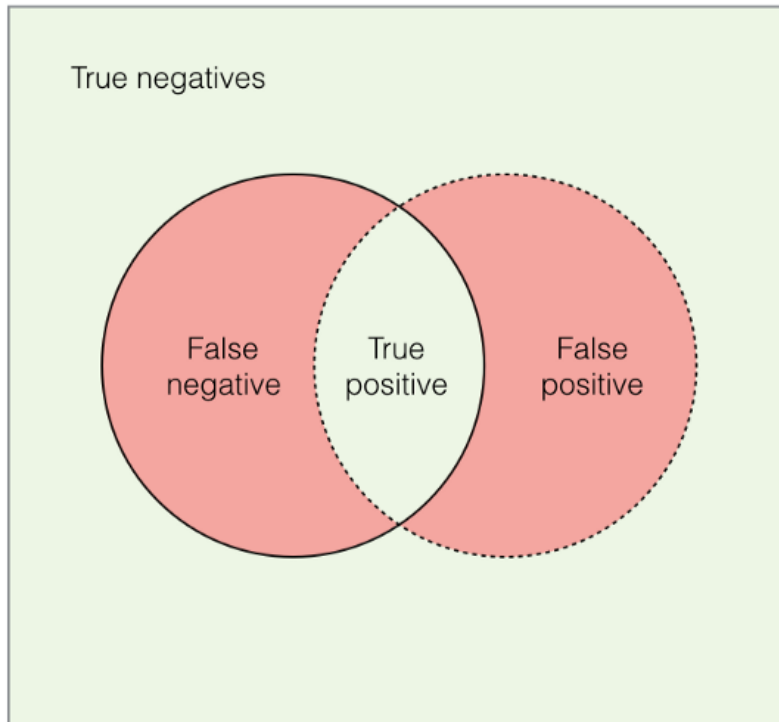
**High Recall, High Precision**



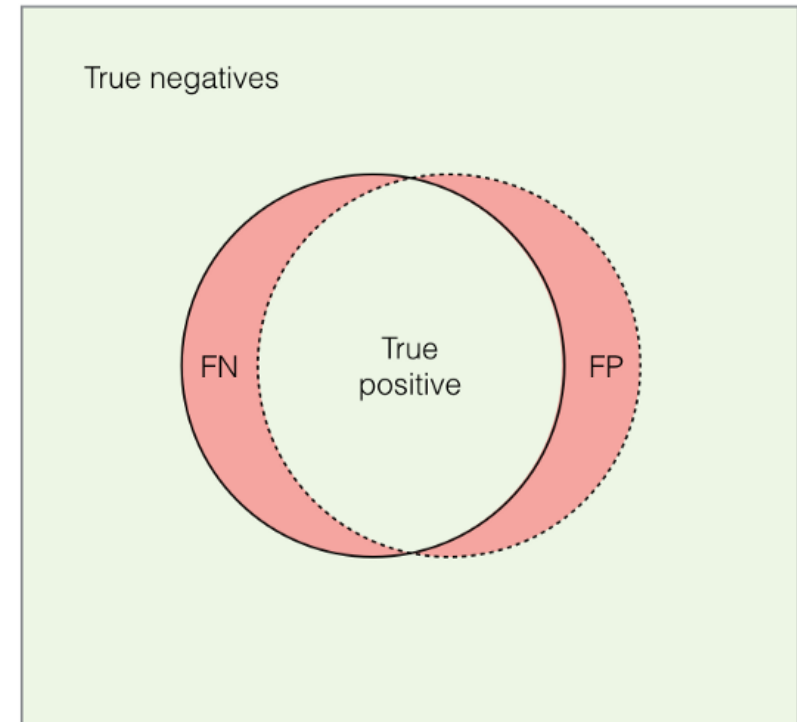
# Precision and Recall

---

**Low Recall, Low Precision**



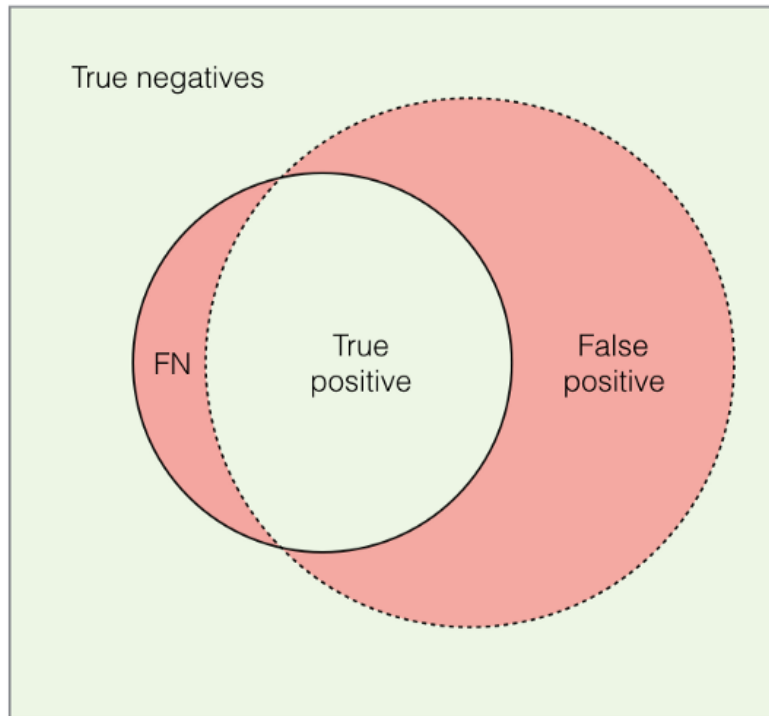
**High Recall, High Precision**



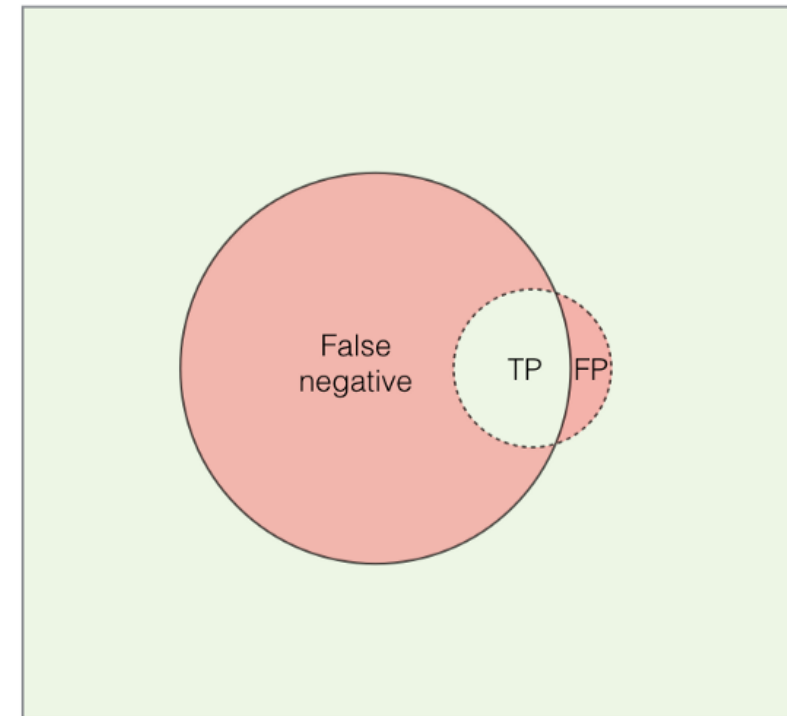
# Precision and Recall

---

**High Recall, Low Precision**

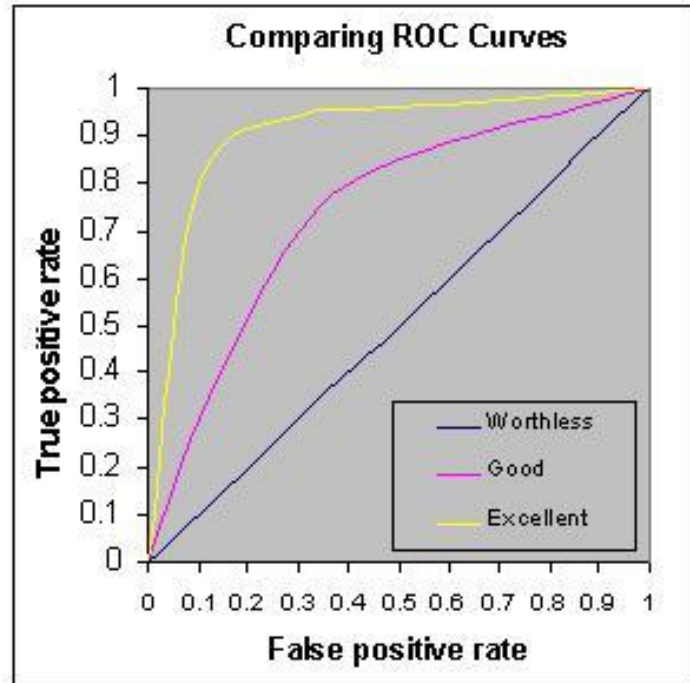


**Low Recall, High Precision**



# Area Under ROC

- ROC: Receiver Operating Characteristic
- Represent model's ability to discriminate between positive and negative classes (for binary classification)



ROC area of 1.0 represents a model that made all predictions perfectly  
ROC area of 0.5 represents a model as good as random

- X-axis: 1 – specificity (false positive rate =  $FP/(FP+TN)$ )
- Y-axis: sensitivity (true positive rate =  $TP/(TP+FN)$ )