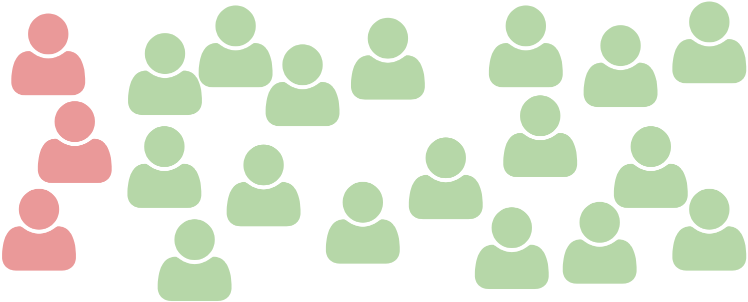
Introduction to Machine Learning

Evaluation: Measures for Binary Classification: ROC Measures

compstat-lmu.github.io/lecture_i2ml

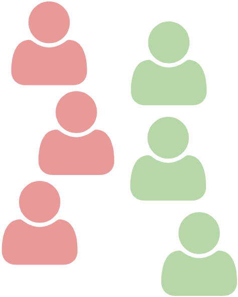
IMBALANCED BINARY LABELS



Classify all as “no disease” (green) $\rightarrow$ high accuracy.

Accuracy Paradox

IMBALANCED COSTS



Classify incorrectly as “no disease” $\rightarrow$ very high cost

CONFUSION MATRIX

		True Class y	
		$+$	$-$
Pred.	$+$	TP	FP
$\hat{y}$	$-$	FN	TN

- $+$: “positive” class
- $-$: “negative” class
- n_+: number of observations in $+$
- n_-: number of observations in $-$

LABELS: ROC METRICS

From the confusion matrix (binary case), we can calculate "ROC" metrics.

		True Class y		
		$+$	$-$	
Pred.	$+$	TP	FP	PPV = $\frac{\text{TP}}{\text{TP}+\text{FP}}$
$\hat{y}$	$-$	FN	TN	NPV = $\frac{\text{TN}}{\text{FN}+\text{TN}}$
		TPR = $\frac{\text{TP}}{\text{TP}+\text{FN}}$	TNR = $\frac{\text{TN}}{\text{FP}+\text{TN}}$	Accuracy = $\frac{\text{TP}+\text{TN}}{\text{TOTAL}}$

- True Positive Rate: How many of the true 1s did we predict as 1?
- True Negative Rate: How many of the true 0s did we predict as 0?
- Positive Predictive Value: If we predict 1 how likely is it a true 1?
- Negative Predictive Value: If we predict 0 how likely is it a true 0?

HISTORY ROC

ROC = receiver operating characteristics

Initially developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battlefields.

http://media.iwm.org.uk/iwm/mediaLib//39/media-39665/large.jpg

Still has the funny name.

LABELS: ROC

Example

		Actual Class y		
		Positive	Negative	
$\hat{y}$ **Pred.**	Positive	**True Positive** (TP) = 20	**False Positive** (FP) = 180	Positive predictive value = TP / (TP + FP) = 20 / (20 + 180) = **10%**
	Negative	**False Negative** (FN) = 10	**True Negative** (TN) = 1820	Negative predictive value = TN / (FN + TN) = 1820 / (10 + 1820) ≈ **99.5%**
		True Positive Rate = TP / (TP + FN) = 20 / (20 + 10) ≈ **67%**	True Negative Rate = TN / (FP + TN) = 1820 / (180 + 1820) = **91%**	

MORE METRICS AND ALTERNATIVE TERMINOLOGY

Unfortunately, for many concepts in ROC, 2-3 different terms exist.

<table>
<tr><td colspan="2"></td><td colspan="2">True condition</td><td colspan="3"></td></tr>
<tr><td></td><td>Total population</td><td>Condition positive</td><td>Condition negative</td><td>Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$</td><td colspan="2">Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$</td></tr>
<tr><td rowspan="2">Predicted condition</td><td>Predicted condition positive</td><td>True positive, Power</td><td>False positive, Type I error</td><td>Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$</td><td colspan="2">False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$</td></tr>
<tr><td>Predicted condition negative</td><td>False negative, Type II error</td><td>True negative</td><td>False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$</td><td colspan="2">Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$</td></tr>
<tr><td colspan="2" rowspan="2"></td><td>True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$</td><td>False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$</td><td>Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$</td><td rowspan="2">Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$</td><td rowspan="2">$F_1$ score = $\frac{1}{\frac{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}{2}}$</td></tr>
<tr><td>False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$</td><td>Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$</td><td>Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$</td></tr>
</table>

▸ Clickable version/picture source

▸ Interactive diagram

LABELS: F_1-MEASURE

A measure that balances two conflicting goals

1. Maximising Positive Predictive Value
2. Maximising True Positive Rate

is the harmonic mean of PPV and TPR:

$$F_1 = 2\frac{PPV \cdot TPR}{PPV + TPR}$$

Note: still doesn't account for the number of true negatives.

LABELS: F_1-MEASURE

Tabulated F_1-Score for different TPR (rows) and PPV (cols) combinations.

	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0	0.00	0.00	0.00	0.00	0.00
0.2	0	0.20	0.27	0.30	0.32	0.33
0.4	0	0.27	0.40	0.48	0.53	0.57
0.6	0	0.30	0.48	0.60	0.69	0.75
0.8	0	0.32	0.53	0.69	0.80	0.89
1.0	0	0.33	0.57	0.75	0.89	1.00

$\rightarrow$ Tends more towards the lower of the 2 combined values.

- $TPR = 0$ or $PPV = 0 \Rightarrow F_1$ of 0
- Predicting always "neg": $F_1 = 0$
- Predicting always "pos": $F_1 = 2PPV/(PPV + 1) = 2n_+/(n_+ + n)$, which will be rather small, if the size of the positive class n_+ is small.