

# Confusion matrix

- ▶ A **confusion matrix** for a problem involving  $n$  classes
  - ▶ is an  $n \times n$  matrix with the columns labeled with actual classes and the rows labels with predicted classes

$$\begin{array}{c} \text{Predicted} \\ \mathbf{Y} \\ \mathbf{N} \end{array} \begin{array}{cc} \mathbf{p} & \mathbf{n} \\ \left( \begin{array}{cc} \text{True positives} & \text{False positives} \\ \text{False negatives} & \text{True negatives} \end{array} \right) \end{array}$$

- ▶ Each example in a test set has an **actual class label** and the **class predicted** by the classifier
- ▶ The confusion matrix separates out the decisions made by the classifier
  - ▶ actual/true classes: **p**(ositive), **n**(egative)
  - ▶ predicted classes: **Y**(es), **N**(o)
  - ▶ The main diagonal contains the count of correct decisions

# Confusion matrix

- Examples

$$CM_A = \begin{array}{c} Y \\ N \end{array} \begin{array}{cc} \text{churn} & \text{not churn} \\ \left( \begin{array}{cc} 500 & 200 \\ 0 & 300 \end{array} \right) \end{array}$$

$$CM_B = \begin{array}{c} Y \\ N \end{array} \begin{array}{cc} \text{churn} & \text{not churn} \\ \left( \begin{array}{cc} 300 & 0 \\ 200 & 500 \end{array} \right) \end{array}$$

# Confusion matrix

- ▶ Based on the entries of the confusion matrix, we can describe various evaluation metrics

- ▶ True positive rate (Recall):  $\frac{TP}{TP+FN}$

- ▶ False negative rate:  $\frac{FN}{TP+FN}$

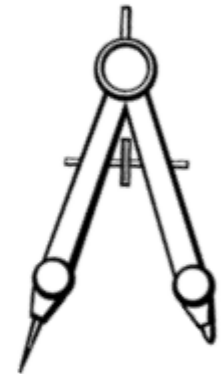
- ▶ Precision (accuracy over the cases predicted to be positive):  $\frac{TP}{TP+FP}$

- ▶ F-measure (harmonic mean):  $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

- ▶ Sensitivity:  $\frac{TN}{TN+FP}$

- ▶ Specificity:  $\frac{TP}{TP+FN}$

- ▶ Accuracy (count of correct decisions):  $\frac{TP+TN}{P+N}$



# Classification weights

- Each outcome “X” (true positive, false positives, etc.) can be associated to a weight
- Objective of the classification algorithm: find the model that minimizes the total cost:
  - $\text{Sum}_x \text{weight}(X) * \text{freq}(X)$

$$CM_A = \begin{matrix} & \text{churn} & \text{not churn} \\ Y & \begin{pmatrix} 500 & 200 \end{pmatrix} \\ N & \begin{pmatrix} 0 & 300 \end{pmatrix} \end{matrix}$$

$$CM_B = \begin{matrix} & \text{churn} & \text{not churn} \\ Y & \begin{pmatrix} 300 & 0 \end{pmatrix} \\ N & \begin{pmatrix} 200 & 500 \end{pmatrix} \end{matrix}$$

# Classification weights

- Example

Outcomes

Predicted	Actual	
	p	n
Y	56	7
N	5	42

Weights

Predicted	Actual	
	p	n
Y	99	-1
N	0	0

- Total cost:  $56*99 + 7*(-1) + 5*0 + 42*0$