Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

#351 Classification

été 2020

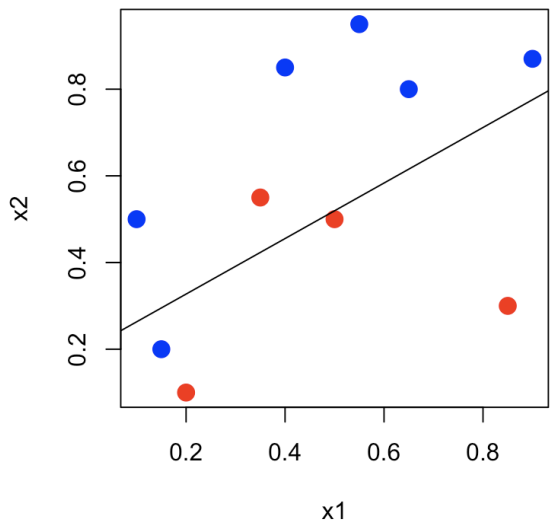
ROC Curve

Given a model, $\widehat{p}(\boldsymbol{x}) = \mathbb{P}Y = 1|\boldsymbol{X} = \boldsymbol{x}$,
consider some threshold $s \in (0, 1)$

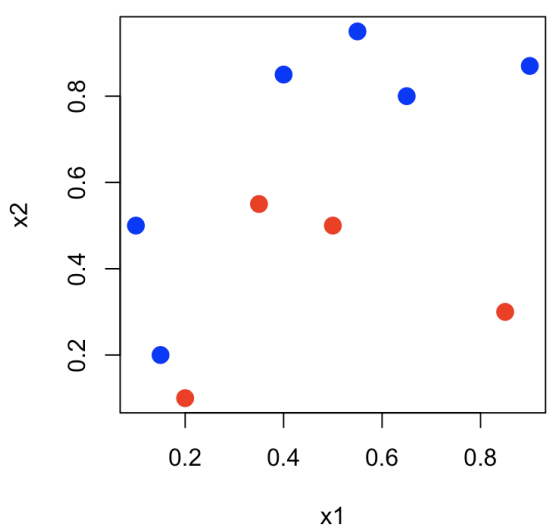
$$\widehat{y}_i = \begin{cases} 1 & \text{if } \widehat{p}_i > \text{threshold } s \\ 0 & \text{if } \widehat{p}_i \leq \text{threshold } s \end{cases}$$

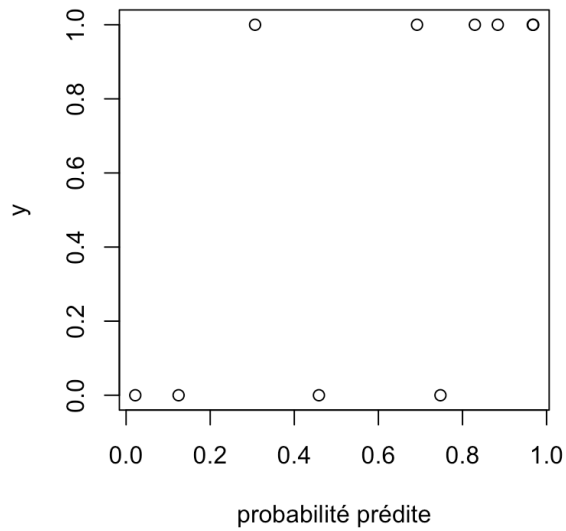
Consider the following confusion matrix

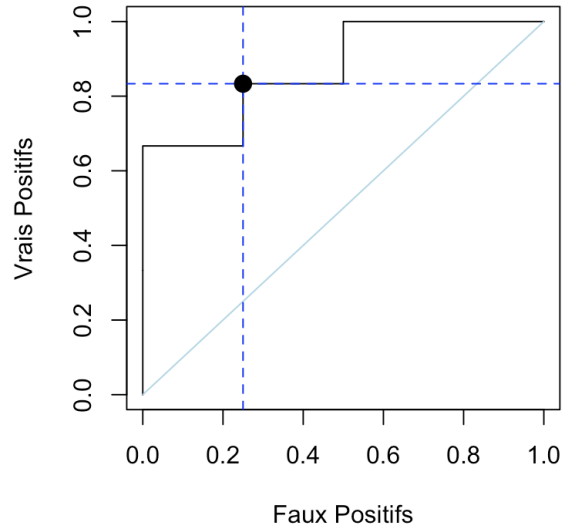
	$\widehat{y} = 0$	$\widehat{y} = 1$	
$y = 0$	3	1	4
$y = 1$	1	5	6
	4	6	10

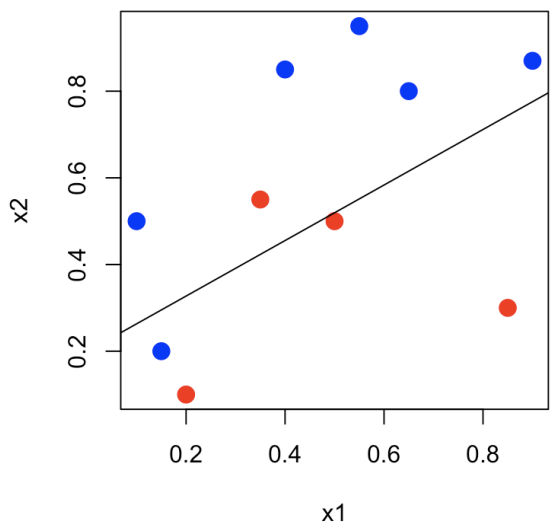


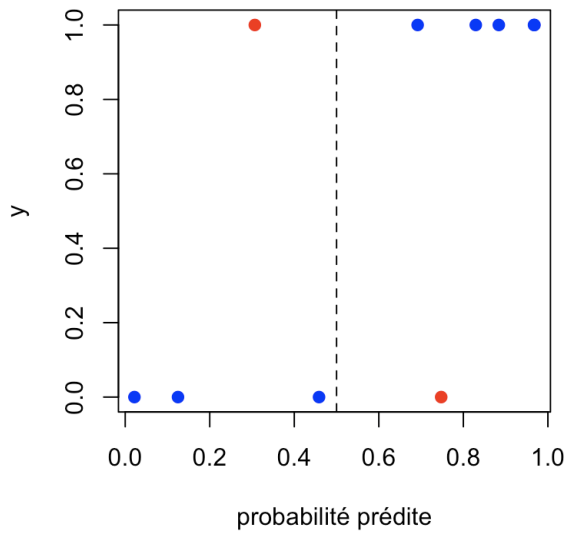
ROC Curve

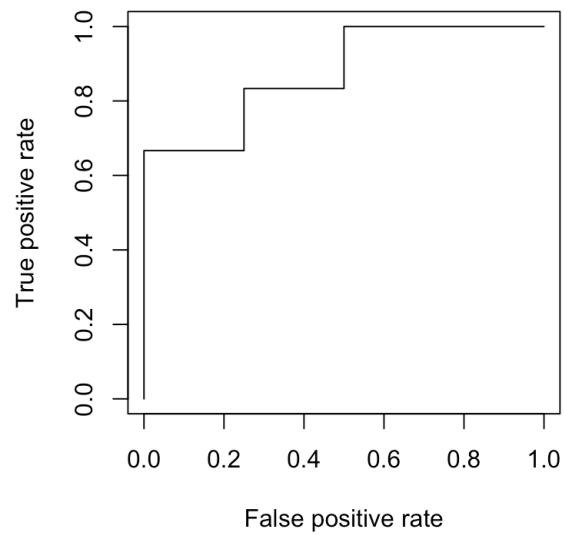

x2
0.2
0.4
0.6
0.8
x1
0.2
0.4
0.6
0.8


y
0.0
0.2
0.4
0.6
0.8
1.0
probabilité prédite
0.0
0.2
0.4
0.6
0.8
1.0


Vrais Positifs
0.0
0.2
0.4
0.6
0.8
1.0
Faux Positifs
0.0
0.2
0.4
0.6
0.8
1.0


x2
0.2
0.4
0.6
0.8
x1
0.2
0.4
0.6
0.8


y
0.0
0.2
0.4
0.6
0.8
1.0
probabilité prédite
0.0
0.2
0.4
0.6
0.8
1.0


True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

ROC Curve

$$FPR = \frac{\mathbb{P}[y = 0, \widehat{y} = 1]}{\mathbb{P}[y = 0]} \text{ et } TPR = \frac{\mathbb{P}[y = 1, \widehat{y} = 1]}{\mathbb{P}[y = 1]}$$