# Machine Learning

## Performance - Naive Bayes

Edgar Francisco Roman-Rangel
edgar.roman@itam.mx

Digital Systems Department.
Instituto Tecnológico Autónomo de México, ITAM.

Diplomado - UAEM

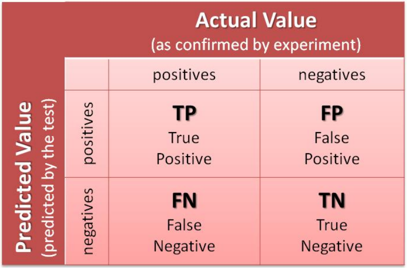
May 1st, 2021.

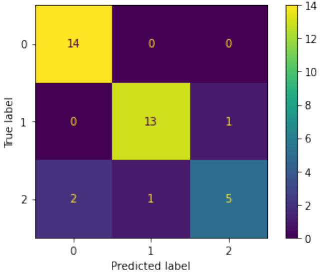
# Outline

Performance metrics

Naive Bayes Classifier

# Confusion matrix



# Other metrics, I

Accuracy:

$$a = \frac{TP + TN}{TP + FP + FN + TN}.$$

Precision:

$$p = \frac{TP}{TP + FP}.$$

Sensitivity (recall):

$$r = \frac{TP}{TP + FN}.$$

Specificity (selectivity):

$$s = \frac{TN}{TN + FP}.$$

# Other metrics, II

False negative rate (FNR):

$$FNR = \frac{FN}{TP + FN}.$$

False positive rate (FPR):

$$FPR = \frac{FP}{FP + TN}.$$

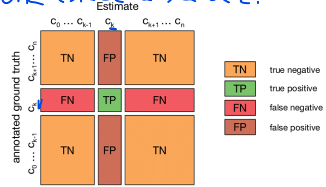
F1 score:

$$F1 = \frac{2TP}{2TP + FP + FN}.$$

# Confusion matrix

Analizando una clase a la vez.

Sólo la k-ésima clase

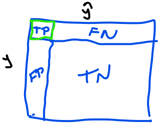



$y$

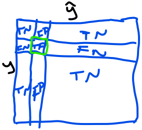
$\hat{y}$

| | Actual Dog | Actual Cat | Actual Rabbit |
|---|---|---|---|
| Classified Dog | 23 | 12 | 7 |
| Classified Cat | 11 | 29 | 13 |
| Classified Rabbit | 4 | 10 | 24 |

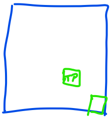
Clase 0




Clase 1




Clase k




Ej)

Dog

TP = 23

TN = 29 + 24 + 13 + 10 = 76

FP = 12 + 7 = 19

FN = 11 + 4 = 15

$$P_D = \frac{TP}{TP+FP} = \frac{23}{23+19} = 0.548$$

Cat

TP = 29

TN = 23 + 7 + 4 + 24 = 58

FP = 11 + 13 = 24

FN = 12 + 10 = 22

$$P_C = \frac{29}{29+24} = 0.547$$

Rabbit

TP = 24

TN = 23 + 12 + 11 + 29 = 75

FP = 4 + 10 = 14

FN = 7 + 13 = 20

$$P_R = \frac{24}{24+14} = 0.632$$

Full DB

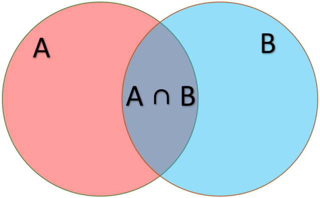
$$P = \frac{23+29+24}{23+29+24+19+24+14} = 0.571$$

# Outline

# Bayes Theorem

$$\underbrace{p(A, B)}_{\text{joint}} = p(A)p(B|A).$$

$p(B|A) = \dfrac{p(A|B)\,p(B)}{p(A)}$




$$p(A)p(B|A) = p(B)p(A|B),$$

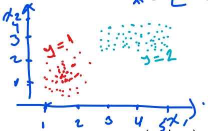
joint

conditional $$p(B|A) = \frac{p(B)p(A|B)}{p(A)}.$$ marginal

# Bayes estimator

Classification method based on the Bayes Theorem:

$\mathbf{x} = [x_1, x_2]$




$p(y=c \mid x=[x_1, .., x_n]) =$

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})},$$

- $p(y|\mathbf{x})$: **posterior** (what we are looking for).
  Probability of class $y$, given input $\mathbf{x}$.
- $p(\mathbf{x}|y)$: **prior** (observation from data).
  Probability of observing input $\mathbf{x}$ when data point is of class $y$.
- $p(y)$: **likelihood**. verosimilitud
  Probability of class $y$ in our data set.
- $p(\mathbf{x})$: **evidence**.
  Probability of input pattern $\mathbf{x}$ in our data set.

## Naive Bayes Classifier

- Assume independence among features (elements of vector $\mathbf{x}$).

$$p(\mathbf{x}) = p(x_1)p(x_2)\dots p(x_N).$$

$p(x) = p(x_1)\,p(x_2|x_1)\,p(x_3|x_2,x_1)\dots$ ← ¡¡NO!! sería si no suponemos independencia.

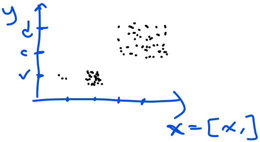
This is rare in real world scenarios. However, it does work in practice.

- Evidence $p(\mathbf{x})$ is the same for a fixed data set.

Therefore, the naive version becomes:

$$p(y|\mathbf{x}) \propto p(\mathbf{x}|y)p(y),$$
$$\propto \prod_{n=1}^{N} p(x_n|y)p(y).$$

# Univariate example




Probability of 'dog' (d) if there are '4 legs'.

$$p(y = \mathsf{d}|4) = \frac{p(4|\mathsf{d})p(\mathsf{d})}{p(4)},$$

suppose from our data we count:

- $p(4|\mathsf{d}) = 4/5$.
- $p(\mathsf{d}) = 2/3$.
- $p(4) = 1/10$.

$$p(y = \mathsf{d}|4) = \frac{4/5 \cdot 2/3}{1/10} = \frac{8/15}{1/10} = 5.3$$

Let's say it is a dog if all other $p(y|4)$ are less than 5.3.

# Multivariate example

Hoy obsemamos $x = [\text{sun}, \text{cool}, T] \to y = ?$

$p(y=0) = 5/14$
$p(y=1) = 9/14$

I love biking. Should I go biking today?

Let us use:

- $\mathbf{x} = [x_1 = \text{sky}, x_2 = \text{temperature}, x_3 = \text{wind}]$.
- $y = \{0, 1\}$ (biking or not biking).

sky

$p(\text{sun}|0) = 3/5$ $\quad p(\text{sun}|1) = 2/9$
$p(\text{cloud}|0) = 0/5$ $\quad p(\text{cloud}|1) = 4/9$
$p(\text{rain}|0) = 2/5$ $\quad p(\text{rain}|1) = 3/9$

temp

$p(\text{hot}|0) = 2/5$ $\quad p(\text{hot}|1) = 2/9$
$p(\text{mild}|0) = 2/5$ $\quad p(\text{mild}|1) = 4/9$
$p(\text{cool}|0) = 1/5$ $\quad p(\text{mild}|1) = 3/9$

wind

$p(T|0) = 3/5$ $\quad p(T|1) = 3/9$
$p(F|0) = 2/5$ $\quad p(F|1) = 6/9$

| n | sky | temp | wind | biking |
|---|---|---|---|---|
| 1 | sunny | hot | FALSE | 0 |
| 2 | sunny | hot | TRUE | 0 |
| 3 | cloudy | hot | FALSE | 1 |
| 4 | rainy | mild | FALSE | 1 |
| 5 | rainy | cool | FALSE | 1 |
| 6 | rainy | cool | TRUE | 0 |
| 7 | cloudy | cool | TRUE | 1 |
| 8 | sky sunny | mild | FALSE | 0 |
| 9 | sky sunny | cool | FALSE | 1 |
| 10 | rainy | mild | FALSE | 1 |
| 11 | sky sunny | mild | TRUE | 1 |
| 12 | cloudy | mild | TRUE | 1 |
| 13 | cloudy | hot | FALSE | 1 |
| 14 | rainy | mild | TRUE | 0 |

$$p(1|x) = p(\text{sun}|1) \cdot p(\text{cool}|1) \cdot p(T|1) \cdot p(1)$$
$$= \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14} = \underline{0.0159}$$

$$p(0|x) = p(\text{sun}|0) \cdot p(\text{cool}|0) \cdot p(T|0) \cdot p(0)$$
$$= \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{5}{14} = \underline{0.0257}$$

$0.0257 > 0.0159$

$\to p(0|x) > p(1|x)$

$\to$ no biking today

# Notes on Naive Bayes Classifier

- This method exploits probabilities.
- Easy and fast.
- Performs better than other methods (assuming independence).
- Zero frequencies might be problematic.

Used for:

- Credit analysis.
- Spam detector.
- Medical analysis.
- Recommendation systems.

## Q&A

Thank you!

edgar.roman@itam.mx