

Machine Learning: Naive Bayes Classifiers

A. Zemmari

LaBRI - Université de Bordeaux

Nov. 28, 2017

An Example and a Bit of Intuition

S(ex)	H(eight) (m)	W(eight) (kg)	F(oot) size (cm)
M	1.82	82	30
M	1.80	86	28
M	1.70	77	30
M	1.80	75	25
F	1.52	45	15
F	1.65	68	20
F	1.68	59	18
F	1.75	68	23

- Question : Is (1.81, 59, 21) male or female?

An Example and a Bit of Intuition

S(ex)	H(eight) (m)	W(eight) (kg)	F(oot) size (cm)
M	1.82	82	30
M	1.80	86	28
M	1.70	77	30
M	1.80	75	25
F	1.52	45	15
F	1.65	68	20
F	1.68	59	18
F	1.75	68	23

- ▶ Question : Is (1.81, 59, 21) male or female?

In other words:

$$\begin{aligned} & \Pr(S = M \mid (H, W, F) = (1.81, 59, 21)) \\ & > \Pr(S = F \mid (H, W, F) = (1.81, 59, 21)) ? \quad (1) \end{aligned}$$

Bayes Law:

$$\Pr(S = M | H, W, F) = \frac{\Pr(S) \times \Pr(H, W, F | S)}{\Pr(H, W, F)}$$

In other words:

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

But $\Pr(H, W, F)$ does not depend on S , so the question boils down to:

$$\begin{aligned} & \Pr(S = M) \times \Pr(H, W, F | S = M) \\ & > \Pr(S = F) \times \Pr(H, W, F | S = F) \end{aligned}$$

Bayes Law:

$$\Pr(S = M | H, W, F) = \frac{\Pr(S) \times \Pr(H, W, F | S)}{\Pr(H, W, F)}$$

In other words:

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

But $\Pr(H, W, F)$ does not depend on S , so the question boils down to:

$$\begin{aligned} & \Pr(S = M) \times \Pr(H, W, F | S = M) \\ & > \Pr(S = F) \times \Pr(H, W, F | S = F) \end{aligned}$$

Bayes Law:

$$\Pr(S = M | H, W, F) = \frac{\Pr(S) \times \Pr(H, W, F | S)}{\Pr(H, W, F)}$$

In other words:

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

But $\Pr(H, W, F)$ does not depend on S , so the question boils down to:

$$\begin{aligned} & \Pr(S = M) \times \Pr(H, W, F | S = M) \\ & > \Pr(S = F) \times \Pr(H, W, F | S = F) \end{aligned}$$

- ▶ $\Pr(S)$ is easy to estimate.
- ▶ What about $\Pr(H, W, F | S)$?

- ▶ $\Pr(S)$ is easy to estimate.
- ▶ What about $\Pr(H, W, F | S)$?

Naive Bayes Hypothesis

We have:

$$\mathbb{P}r(H, W, F | S) = \mathbb{P}r(H | S) \times \mathbb{P}r(W | S, H) \times \mathbb{P}r(F | S, W, H)$$

Naive Bayes:

$$\mathbb{P}r(W | S, H) = \mathbb{P}r(W | S) \text{ et } \mathbb{P}r(F | S, W, H) = \mathbb{P}r(F | S)$$

Naive Bayes Hypothesis

We have:

$$\mathbb{P}r(H, W, F | S) = \mathbb{P}r(H | S) \times \mathbb{P}r(W | S, H) \times \mathbb{P}r(F | S, W, H)$$

Naive Bayes:

$$\mathbb{P}r(W | S, H) = \mathbb{P}r(W | S) \text{ et } \mathbb{P}r(F | S, W, H) = \mathbb{P}r(F | S)$$

Naive Bayes Hypothesis

$$\Pr(W | S, H) = \Pr(W | S)$$

what does that mean?

"Among male individuals, the weight is independent of the height"

What do you think?

Despite that naive assumption, Naive Bayes classifiers perform very well !

Let's formalize that a little more.

Naive Bayes Hypothesis

$$\Pr(W | S, H) = \Pr(W | S)$$

what does that mean?

"Among male individuals, the weight is independent of the height"

What do you think?

Despite that naive assumption, Naive Bayes classifiers perform very well !

Let's formalize that a little more.

Naive Bayes Hypothesis

$$\Pr(W | S, H) = \Pr(W | S)$$

what does that mean?

"Among male individuals, the weight is independent of the height"

What do you think?

Despite that naive assumption, Naive Bayes classifiers perform very well !

Let's formalize that a little more.

Naive Bayes Hypothesis

$$\Pr(W | S, H) = \Pr(W | S)$$

what does that mean?

"Among male individuals, the weight is independent of the height"

What do you think?

Despite that naive assumption, Naive Bayes classifiers perform very well !

Let's formalize that a little more.

Naive Bayes Hypothesis

$$\Pr(W | S, H) = \Pr(W | S)$$

what does that mean?

"Among male individuals, the weight is independent of the height"

What do you think?

Despite that naive assumption, Naive Bayes classifiers perform very well !

Let's formalize that a little more.

Naive Bayes classifiers in one slide!

$$\mathbb{P}r(Y | X_1, \dots, X_n) = \frac{\mathbb{P}r(Y) \times \mathbb{P}r(X_1, \dots, X_n | Y)}{\mathbb{P}r(X_1, \dots, X_n)}$$

Naive conditional independence assump.:

$$\forall i \neq j, \mathbb{P}r(X_i | Y, X_j) = \mathbb{P}r(X_i | Y)$$

$$\Rightarrow \mathbb{P}r(Y | X_1, \dots, X_n) = \frac{1}{Z} \times \mathbb{P}r(Y) \times \prod_{i=1}^n \mathbb{P}r(X_i | Y)$$

Naive Bayes classifiers in one slide!

$$\mathbb{P}r(Y | X_1, \dots, X_n) = \frac{\mathbb{P}r(Y) \times \mathbb{P}r(X_1, \dots, X_n | Y)}{\mathbb{P}r(X_1, \dots, X_n)}$$

Naive conditional independence assump.:

$$\forall i \neq j, \mathbb{P}r(X_i | Y, X_j) = \mathbb{P}r(X_i | Y)$$

$$\Rightarrow \mathbb{P}r(Y | X_1, \dots, X_n) = \frac{1}{Z} \times \mathbb{P}r(Y) \times \prod_{i=1}^n \mathbb{P}r(X_i | Y)$$

Naive Bayes classifiers in one slide!

$$\mathbb{P}r(Y | X_1, \dots, X_n) = \frac{\mathbb{P}r(Y) \times \mathbb{P}r(X_1, \dots, X_n | Y)}{\mathbb{P}r(X_1, \dots, X_n)}$$

Naive conditional independence assump.:

$$\forall i \neq j, \mathbb{P}r(X_i | Y, X_j) = \mathbb{P}r(X_i | Y)$$

$$\Rightarrow \mathbb{P}r(Y | X_1, \dots, X_n) = \frac{1}{Z} \times \mathbb{P}r(Y) \times \prod_{i=1}^n \mathbb{P}r(X_i | Y)$$

Back to the example

$$\Pr(S | H, W, F) = \frac{1}{Z} \times \Pr(S) \times \Pr(H | S) \times \Pr(W | S) \times \Pr(F | S)$$

S(ex)	H(eight) (m)	W(eight) (kg)	F(oot) size (cm)
M	1.82	82	30
M	1.80	86	28
M	1.70	77	30
M	1.80	75	25
F	1.52	45	15
F	1.65	68	20
F	1.68	59	18
F	1.75	68	23

Back to the example

$$\Pr(S = M) =? \quad (1)$$

$$\Pr(H = 1.81 \mid S = M) =?$$

$$\Pr(W = 59 \mid S = M) =?$$

$$\Pr(F = 21 \mid S = M) =?$$

Back to the example

$$\mathbb{P}r(S | H, W, F) = \frac{1}{Z} \times \mathbb{P}r(S) \times \mathbb{P}r(H | S) \times \mathbb{P}r(W | S) \times \mathbb{P}r(F | S)$$

S is discrete, H, W and F are assumed Gaussian.

S	\hat{p}_S	$\hat{\mu}_{H S}$	$\hat{\sigma}_{H S}$	$\hat{\mu}_{W S}$	$\hat{\sigma}_{W S}$	$\hat{\mu}_{F S}$	$\hat{\sigma}_{F S}$
M	0.5	1.78	0.0469	80	4.3012	28.25	2.0463
F	0.5	1.65	0.0834	60	9.4074	19	2.9154

$$\begin{aligned}\mathbb{P}r(S = M | 1.81, 59, 21) &= \frac{1}{Z} \times 0.5 \times \frac{e^{-\frac{(1.78-1.81)^2}{2 \cdot 0.0469^2}}}{\sqrt{2\pi} \cdot 0.0469} \times \frac{e^{-\frac{(80-59)^2}{2 \cdot 4.3012^2}}}{\sqrt{2\pi} \cdot 4.3012} \times \frac{e^{-\frac{(28.25-21)^2}{2 \cdot 2.0463^2}}}{\sqrt{2\pi} \cdot 2.0463} \\ &= \frac{1}{Z} \times 7.854 \cdot 10^{-10}\end{aligned}$$

$$\mathbb{P}r(S = F | 1.81, 59, 21) = \frac{1}{Z} \times 1.730 \cdot 10^{-3}$$

Back to the example

$$\mathbb{P}r(S | H, W, F) = \frac{1}{Z} \times \mathbb{P}r(S) \times \mathbb{P}r(H | S) \times \mathbb{P}r(W | S) \times \mathbb{P}r(F | S)$$

Conclusion: given the data, (1.81m, 59kg, 21cm) is more likely to be female.

Features

$$\mathbb{P}r(Y | X_1, \dots, X_n) = \frac{1}{Z} \times \mathbb{P}r(Y) \times \prod_{i=1}^n \mathbb{P}r(X_i | Y)$$

- ▶ Continuous X_i :
Assume normal distribution $\rightarrow X_i | Y = y \sim \mathcal{N}(\mu_{iy}, \sigma_{iy})$
- ▶ Binary X_i :
Assume Bernoulli distribution $\rightarrow X_i | Y = y \sim \mathcal{B}(p_{iy})$

Algorithm

▶ Train:

For all possible values of Y and X_i , compute $\hat{\mathbb{P}}_r(Y = y)$ and $\hat{\mathbb{P}}_r(X_i = x_i | Y = y)$

▶ Predict:

Given (x_1, x_2, \dots, x_n) , return y that maximizes $\hat{\mathbb{P}}_r(Y = y) \hat{\mathbb{P}}_r(X_i = x_i | Y = y)$