Machine Learning (Lab support)

Naïve Bayes

Abdelkrime Aries

Laboratoire de la Communication dans les Systèmes Informatiques (LCSI) École nationale Supérieure d'Informatique (ESI, ex. INI), Algiers, Algeria

Academic year: 2024-2025







License: CC-BY 4.0



Attribution 4.0 International (CC BY 4.0)

https://creativecommons.org/licenses/by/4.0/deed.en

You are free to:

Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material for any purpose, even commercially.



The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

 $\begin{tabular}{lll} \textbf{No additional restrictions} & -- You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits. \end{tabular}$

Numerical application

Theoretical formulation

Machine Learning (Lab support)

Naïve Bayes: Bayes theory

Posterior
$$P(A|B) = \frac{P(A) P(B|A)}{P(B|A)}$$
Evidence

ESI.ML.Lab Aries (2024-2025)

Theoretical formulation

Machine Learning (Lab support)

Naïve Bayes: Plan

- Theoretical formulation
 - Estimation
 - Prior probability
 - Likelihood

- 2
 - Numerical application
 - Multinomial NB
 - Bernoulli NB
 - Normal NB

Theoretical formulation
Numerical application

Estimation Prior probability Likelihood

Section 1

Theoretical formulation

1011	ر	l

Estimation

Likelihood

Prior probability

Numerical application

Theoretical formulation

Given a sample x, the probability of generating a class k can be expressed as:

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

 Given L classes, the output class is the one that maximizes this probability

$$\hat{y} = \arg\max_{x} p(y = k|x), \ k \in \{1, \dots, L\}$$

• in this case, no need for Evidence probability (not dependent to y) $p(y = k|x) \propto p(y = k)p(x|y = k)$

Naïve Bayes

Theoretical formulation

	,	•	•	
ıt	İ	C	כ	ľ
			1	

Estimation

Likelihood

Naïve Bayes: Theoretical formula

Theoretical formulation Numerical application

• The output class \hat{y} is estimated as $\hat{y} = \arg\max p(y = k|x) = \arg\max p(y = k)p(x|y = k), \ k \in \{1, \dots, L\}$

The naive part: the assumption of features independence

$$p(x|y=k) \approx \prod_{j=1} P(x_j|y=k)$$

In this case, the estimation function would be:

$$\hat{y} = \arg\max_{y=k} p(y=k) \prod_{j=1} P(x_j | y=k), \ k \in \{1, \dots, L\}$$

• In practice, the calculation is simplified
$$\hat{y} = \arg\max_{y=k} \log p(y=k) + \sum_{i=1}^{N} \log p(x_{j}|y=k), \ k \in \{1, \cdots, L\}$$

Estimation

Theoretical formulation

Estimation

Likelihood

Prior probability

Naïve Bayes: Theoretical formulation Prior probability

$$p(y = k) = \frac{|\{y^{(i)} = k, \ i \in \{1, \dots, M\}\}|}{M}$$

- $|\{y^{(i)} = k, i \in \{1, \dots, M\}\}|$ is the number of training samples having k as class
- M is the size of the training dataset
- If classes' distribution is uniform, this probability can be ignored
- If we want to give the same prior probability to classes, this probability can be ignored

Estimation

Likelihood

Prior probability

Theoretical formulation

Numerical application

$$p(x_j = v | y = k) = \frac{|\{y^{(i)} = k \land x_j^{(i)} = v, \ i \in \{1, \dots, M\}\}|}{|\{y^{(i)} = k, \ i \in \{1, \dots, M\}\}|} = \frac{\#(y = k \land x_j = v)}{\#(y = k)}$$

- x_i is a categorical feature having a value v
- v is a value among unique values V_i (called vocabulary) of the feature j
- $\#(y = k \land x_i = v)$ is the number of training samples with feature j equals
- to v and having k as class
- #(y = k) is the number of training samples having k as class • Smoothing can be used in case there are unseen values v in the test

dataset, where
$$V_j$$
 is the vocabulary of the feature j (unique categories)
$$P(x_j = v|y_k) = \frac{\#(y = k \land x_j = v) + \alpha}{\#(y = k) + \alpha|V_j|}$$

sklearn.naive_bayes.CategoricalNB



 $p(word = w_j | y = k) = \frac{C_{jk} + \alpha}{C_{\iota} + \alpha |V|}$

Estimation

Likelihood

Prior probability

Naïve Bayes: Theoretical formulation

Likelihood: Multinomial distribution (text)

$$\hat{y} = \arg\max_{k} p(y = k) * \prod_{w \in Lext} p(word = w|y = k)$$

Theoretical formulation

Numerical application

- A text can be seen as one feature with words as values
- C_k is the number of training samples having k as class
- C_{ik} is the number of occurrences of word w_i in texts having k as class
- V is the vocabulary (unique words in the training dataset)
- sklearn.naive_bayes.MultinomialNB

 $p(x_i = v|y = k) = p(x_i = 1|y = k)v + (1 - p(x_i = 1|y = k))(1 - v)$

Estimation

Likelihood

Prior probability

Theoretical formulation

 $p(x_j = 1|y_k) = \frac{|\{x_j^{(i)} = 1 \land y^{(i)} = k, \ i \in \{1, \dots, M\}\}|}{|\{y^{(i)} = k, \ i \in \{1, \dots, M\}\}|}$

•
$$x_i$$
 is a boolean feature having a value $v \in \{0, 1\}$

sklearn.naive_bayes.BernoulliNB

_)	ļ		
t	(0	ľ	1

Estimation

Likelihood

Prior probability

Naïve Bayes: Theoretical formula Likelihood: Normal (Gaussian) distribution

$$p(x_j = v|Y_k) = \frac{1}{\sqrt{2\pi\sigma_{kj}^2}} e^{\frac{-(v-\mu_{kj})^2}{2\sigma_{kj}^2}}$$

Theoretical formulation

Numerical application

- x_i is a numerical feature having values $v \in]-\infty, +\infty[$
- μ_{ki} is the mean of x_i 's values having k as class
- σ_{ki}^2 is the **unbiased** variance of x_j 's values having k as class

sklearn.naive_bayes.GaussianNB

This slide is left blank

for no reason

Theoretical formulation
Numerical application

Bernoulli NB Normal NB

Section 2

Numerical application

14/21

ESI.ML.Lab Aries (2024-2025)

ML-Lab: Naïve Bayes

Multinomial NB
Bernoulli NB
Normal NB

Naïve Bayes

Numerical application

- Multinomial NB: Play or not based on these categorical features: outlook (sunny, overcast, rainy), temp (hot, mild, cool), humidity (high, normal), windy (true, false)
- Bernoulli NB: Pass the exam or fail based on these boolean features: confident, studied, sick
- Normal NB: Male or female based on these numerical features: height (cm), weight (kg), footsize (cm)

Theoretical formulation Numerical application

Bernoulli NB Normal NB

Multinomial NB

Naïve Bayes: Numerical application

Multinomial NB: Example (1)

temp humidity windy play outlook

hot

- hiah false nο high true no
- sunnv hot overcast hot high false ves rainy mild high false ves
- false rainv cool normal ves rainy cool normal true no
- overcast cool normal true ves mild false sunnv hiah no false sunnv cool normal yes
- rainy mild normal false ves sunny mild normal true ves
- overcast mild hiah true overcast hot normal false

mild

high

- Prior probability
 - $p(play = yes) = \frac{\#(play = yes)}{M} = \frac{9}{14}$
- $p(play = no) = \frac{\#(play = no)}{M} = \frac{5}{14}$ • Likelihood probability of *outlook* = *rainy*
 - p(outlook = rainy|play = yes) =
 - $\frac{\#(outlook=rainy \land play=yes)}{\#(play=yes)} = \frac{3}{9}$ • p(outlook = rainy|play = no) =
 - $\frac{\#(outlook=rainy \land play=no)}{\#(play=yes)} = \frac{2}{5}$
- Likelihood probability of temp = hot
 - p(temp = hot|play = yes) =
 - $\frac{\#(temp=hot \land play=yes)}{\#(play=yes)} = \frac{2}{9}$ p(temp = hot|play = no) = $\frac{\#(temp=hot \land play=no)}{\#(play=yes)} = \frac{2}{5}$

rainy

sunnv

ves

ves

no

true

Theoretical formulation	
Numerical application	

Normal NB

Multinomial NB Bernoulli NB

Naïve Bayes: Numerical application Multinomial NB: Example (2)

Likelihood probability of humidity = high

Likelihood probability of windy = false

- $p(humidity = high|play = yes) = \frac{\#(humidity = high \land play = yes)}{\#(play = yes)} = \frac{3}{9}$
- $p(humidity = high|play = no) = \frac{\#(humidity = high \land play = no)}{\#(play = ves)} = \frac{4}{5}$
- $p(windy = false|play = yes) = \frac{\#(windy = false \land play = yes)}{\#(play = yes)} = \frac{6}{9}$ • $p(windy = false|play = no) = \frac{\#(windy = false \land play = no)}{\#(play = ves)} = \frac{2}{5}$

Given
$$\vec{v} = [rainy, hot, high, false]$$

- $p(play = yes|x = \vec{v}) \propto \frac{9}{14} (\frac{3}{9} \frac{2}{9} \frac{3}{9} \frac{6}{9}) = \frac{6}{567} \approx 0.0106$
- $p(play = no|x = \vec{v}) \propto \frac{5}{14} (\frac{2}{5} + \frac{2}{5} + \frac{4}{5}) = \frac{16}{875} \approx 0.0183$
- $\hat{\mathbf{v}} = no$

Normal NB

Multinomial NB

Naïve Bayes: Numerical application Bernoulli NB: Example (1)

confident studied sick result fail pass fail pass pass

Prior probability • $p(pass) = \frac{\#(pass)}{M} = \frac{3}{5}$

•
$$p(pass) = \frac{1}{M} = \frac{1}{5}$$

• $p(fail) = \frac{\#(fail)}{M} = \frac{2}{5}$

Probability
$$p(confident|pass) = \frac{2}{3}$$

•
$$p(confident|fail) = \frac{1}{2}$$

•
$$p(studied|pass) = \frac{2}{3}$$

•
$$p(studied|fail) = \frac{1}{2}$$

• $p(sick|fail) = \frac{1}{2}$

$$p(sick|fail) = \frac{1}{2}$$

Given
$$\vec{v} = [1, 0, 0]$$

 $\hat{\mathbf{v}} = fail$

$$p(sick|pass) = \frac{2}{3}$$

$$(sick|fuii) - \frac{1}{2}$$

•
$$p(pass|\vec{v}) \propto \frac{3}{5} \left[\frac{2}{3} (1 - \frac{2}{3})(1 - \frac{2}{3})\right] = \frac{2}{45} \approx 0.0444$$

$$p(sick|pass) = \frac{2}{3}$$

$$, 0, 0]$$

•
$$p(pass|\vec{v}) \propto \frac{3}{5} \left[\frac{2}{3}(1 - \frac{2}{3})(1 - \frac{2}{3})\right] = \frac{2}{45} \approx 0.04$$

• $p(fail|\vec{v}) \propto \frac{2}{5} \left[\frac{1}{2}(1 - \frac{1}{2})(1 - \frac{1}{2})\right] = \frac{1}{20} \approx 0.05$

Theoretical formulation
Numerical application

person

Normal NR

Prior probability: no need since the classes

weight

Multinomial NB

Bernoulli NB

Naïve Bayes: Numerical application Normal NB: Example

height weight footsize person

182

152

168

165

175

19/21

180	86.2	28	male
170	77.1	30	male
180	74.8	25	mala

45.4 15

68.0 23

20

68.0

59.0

81.6 30

male

male

female

ESI.ML.Lab Aries (2024-2025)

 σ^2 μ σ^2 male 178 29.33 79.92 25.48 28.25 5.58 114.04 female 165 92.67 60.1

distribution is uniform

height

Given $\vec{v} = [183, 59, 20]$

$$|mal_{e}\rangle = \frac{1}{2e^{2}} e^{\frac{-(183-178)^{2}}{2e^{2}9}} \approx 0.0481017$$

•
$$p(height = 183|male) = \frac{1}{\sqrt{2\pi * 29.33}} e^{\frac{-(183-178)^2}{2*29.33}} \approx 0.04810173$$

•
$$p(height = 183|female) = \frac{1}{\sqrt{2\pi*92.67}}e^{\frac{-(183-165)^2}{2*92.67}} \approx 0.00721463$$

footsize

11.33

19

Section 3 **Bibliography** 20/21 ESI.ML.Lab Aries (2024-2025) ML-Lab: Naïve Bayes

Bibliography

Bibliography

Bibliography



Metsis, V., Androutsopoulos, I., and Paliouras, G. (2006). Spam filtering with naive bayes - which naive bayes? In International Conference on Email and Anti-Spam.

The probability of finding another slide given the slides you already seen İS