92586 Computational Linguistics

Lesson 3. Naïve Bayes

Alberto Barrón-Cedeño

Alma Mater Studiorum-Università di Bologna
a.barron@unibo.it @_albarron_

03/03/2020

Previously

- Pre-processing (e.g., tokenisation, stemming, stopwording)

Previously

- Pre-processing (e.g., tokenisation, stemming, stopwording)
- BoW representation

Previously

- Pre-processing (e.g., tokenisation, stemming, stopwording)
- BoW representation
- One rule-based sentiment analyser

Table of Contents

Naïve Bayes

Naïve Bayes

Naïve Bayes

1. Introduced in the IR community by Maron (1961)

Naïve Bayes

1. Introduced in the IR community by Maron (1961)
2. First machine-learning approach

Naïve Bayes

1. Introduced in the IR community by Maron (1961)
2. First machine-learning approach
3. It is a supervised model

Naïve Bayes

1. Introduced in the IR community by Maron (1961)
2. First machine-learning approach
3. It is a supervised model
4. It applies Bayes' theorem with strong (naïve) independence assumptions between the features (they are independent; they contribute the same)

Naïve Bayes

1. Introduced in the IR community by Maron (1961)
2. First machine-learning approach
3. It is a supervised model
4. It applies Bayes' theorem with strong (naïve) independence assumptions between the features (they are independent; they contribute the same)

Let us take it easy this time

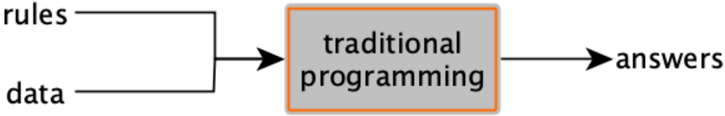
Machine Learning

"the scientific study of algorithms and statistical models that computer systems use to perform a specific task **without using explicit instructions**, relying on patterns and inference instead"

Machine Learning

A change of paradigm

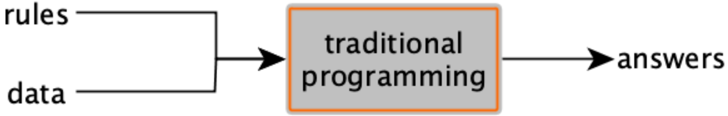
From hand-crafted rules.



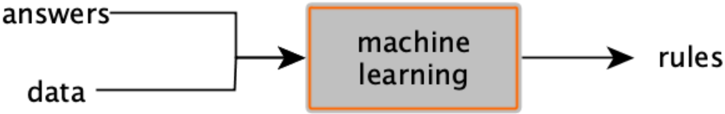
Machine Learning

A change of paradigm

From hand-crafted rules.



To training



Diagrams borrowed from L. Moroney's Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning

Supervised vs Unsupervised

Supervised The algorithms build a mathematical model of a set of data that contains **both the inputs and the desired outputs**

Supervised vs Unsupervised

Supervised The algorithms build a mathematical model of a set of data that contains **both the inputs and the desired outputs**

Unsupervised The algorithms take a set of data that contains **only inputs**, and find structure in the data

https://en.wikipedia.org/wiki/Machine_learning

Naïve Bayes

A conditional probability model

Given an instance represented by a vector

$$\mathbf{x} = (x_1, \ldots, x_n) \tag{1}$$

representing n **independent** features

From https://en.wikipedia.org/wiki/Naive_Bayes_classifier

Naïve Bayes

A conditional probability model

Given an instance represented by a vector

$$\mathbf{x} = (x_1, \dots, x_n) \tag{1}$$

representing n **independent** features

The model assigns the instance the probability

$$p(C_k \mid x_1, \dots, x_n) \tag{2}$$

for each of the k possible outcomes C_k

From `https://en.wikipedia.org/wiki/Naive_Bayes_classifier`

Naïve Bayes

A conditional probability model

Given an instance represented by a vector

$$\mathbf{x} = (x_1, \ldots, x_n) \quad (1)$$

representing n **independent** features

The model assigns the instance the probability

$$p(C_k \mid x_1, \ldots, x_n) \quad (2)$$

for each of the k possible outcomes C_k

From `https://en.wikipedia.org/wiki/Naive_Bayes_classifier`

Naïve Bayes'

Using Bayes' Theorem

The conditional probability $p(C_k \mid x_1, \ldots, x_n)$ can be decomposed as

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x} \mid C_k)}{p(\mathbf{x})} \tag{3}$$

From https://en.wikipedia.org/wiki/Naive_Bayes_classifier

Naïve Bayes'

Using Bayes' Theorem

The conditional probability $p(C_k \mid x_1, \ldots, x_n)$ can be decomposed as

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x} \mid C_k)}{p(\mathbf{x})} \tag{3}$$

How to read this

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

From https://en.wikipedia.org/wiki/Naive_Bayes_classifier

Naïve Bayes'

Using Bayes' Theorem

The conditional probability $p(C_k \mid x_1, \ldots, x_n)$ can be decomposed as

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x} \mid C_k)}{p(\mathbf{x})} \tag{3}$$

How to read this

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

But $p(x)$ does not depend on the class (it's constant!):

$$p(C_k \mid \mathbf{x}) = p(C_k)\ p(\mathbf{x} \mid C_k) \tag{4}$$

From `https://en.wikipedia.org/wiki/Naive_Bayes_classifier`

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{5}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

[1]Symbol | means "given": the probability of the class given the representation vector

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{5}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

$p(C \mid \mathbf{x})$ Posterior probability of the class given the input[1]

[1] Symbol | means "given": the probability of the class given the representation vector

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{5}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

$p(C \mid \mathbf{x})$ Posterior probability of the class given the input[1]

```
if p > 0.5:
    class = positive
else:
    class = negative
```

[1]Symbol | means "given": the probability of the class given the representation vector

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{6}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{6}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

$p(C)$ Prior probability of the class

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{6}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

$p(C)$ Prior probability of the class
How many of the instances I have seen in the past (during training) are positive?

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{7}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{7}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

$p(\mathbf{x} \mid C)$ Prior probability of the class

Naïve Bayes

Going deeper (assuming a binary classifier)

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{7}$$

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}}$$

$p(\mathbf{x} \mid C)$ Prior probability of the class
The probability of the document given the class

Rough Idea

- The value of a particular feature is **independent** of the value of any other feature, given the class variable

From (Lane et al., 2019, p. 65–68)

Rough Idea

- The value of a particular feature is **independent** of the value of any other feature, given the class variable
- All features contribute the same to the classification

From (Lane et al., 2019, p. 65–68)

Rough Idea

- The value of a particular feature is **independent** of the value of any other feature, given the class variable
- All features contribute the same to the classification

- It tries to find keywords in a set of documents that are predictive of the target (output) variable

From (Lane et al., 2019, p. 65–68)

Rough Idea

- The value of a particular feature is **independent** of the value of any other feature, given the class variable
- All features contribute the same to the classification

- It tries to find keywords in a set of documents that are predictive of the target (output) variable
- The internal coefficients will try to map tokens to scores

From (Lane et al., 2019, p. 65–68)

Rough Idea

- The value of a particular feature is **independent** of the value of any other feature, given the class variable
- All features contribute the same to the classification

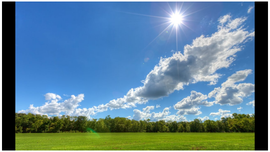
- It tries to find keywords in a set of documents that are predictive of the target (output) variable
- The internal coefficients will try to map tokens to scores
- Same as VADER, but this time **the machine will find the best scores!**

From (Lane et al., 2019, p. 65–68)

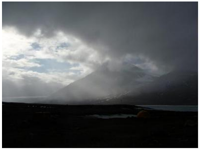
Naïve Bayes

A toy example: Should I play *calcio* today?

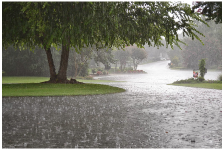
One single factor: weather



sunny



overcast



rainy

(get ready for some of the densest slides I have ever made!)

Naïve Bayes

A toy example: Should I play *calcio* today?

Dataset

Outlook	**Play**
sunny	yes
overcast	yes
sunny	no
rainy	yes
sunny	yes
overcast	yes
sunny	yes
overcast	yes
rainy	yes
sunny	no
rainy	no
overcast	yes
rainy	no
rainy	no

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Dataset

Outlook	**Play**
sunny	yes
overcast	yes
sunny	no
rainy	yes
sunny	yes
overcast	yes
sunny	yes
overcast	yes
rainy	yes
sunny	no
rainy	no
overcast	yes
rainy	no
rainy	no

Computing **all** the probabilities by "counting"

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Dataset

Outlook	Play
sunny	yes
overcast	yes
sunny	no
rainy	yes
sunny	yes
overcast	yes
sunny	yes
overcast	yes
rainy	yes
sunny	no
rainy	no
overcast	yes
rainy	no
rainy	no

Computing **all** the probabilities by "counting"

Frequency table

	Play	
Outlook	**yes**	**no**
sunny	3	2
overcast	4	0
rainy	2	3

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Dataset

Outlook	Play
sunny	yes
overcast	yes
sunny	no
rainy	yes
sunny	yes
overcast	yes
sunny	yes
overcast	yes
rainy	yes
sunny	no
rainy	no
overcast	yes
rainy	no
rainy	no

Computing **all** the probabilities by "counting"

Frequency table

	Play	
Outlook	**yes**	**no**
sunny	3	2
overcast	4	0
rainy	2	3

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	$3/9$	$2/5$
overcast	$4/9$	$0/5$
rainy	$2/9$	$3/5$

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14	5/14

Adapted from http://www.saedsayad.com/naive_bayesian.htm

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14	5/14

[1] $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

Adapted from http://www.saedsayad.com/naive_bayesian.htm

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14 [2]	5/14

[1] $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

[2] $p(c) = p(\text{yes}) = 9/14 = 0.64$

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14 [2]	5/14

[1] $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

[2]
$$p(c) = p(\text{yes}) = 9/14 = 0.64$$
$$p(x) = p(\text{sunny}) = 5/14 = 0.36$$

Adapted from http://www.saedsayad.com/naive_bayesian.htm

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14 [2]	5/14

1 $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

2
$$p(c) = p(\text{yes}) = 9/14 = 0.64$$
$$p(x) = p(\text{sunny}) = 5/14 = 0.36$$

What is the Naïve Bayes' probability of playing if it's **sunny**?

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14 [2]	5/14

[1] $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

[2] $p(c) = p(\text{yes}) = 9/14 = 0.64$

$p(x) = p(\text{sunny}) = 5/14 = 0.36$

What is the Naïve Bayes' probability of playing if it's **sunny**?

$$p(c \mid x) \quad = \quad p(c)p(x \mid c)/p(x)$$

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14 [2]	5/14

[1] $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

[2]
$$p(c) = p(\text{yes}) = 9/14 = 0.64$$
$$p(x) = p(\text{sunny}) = 5/14 = 0.36$$

What is the Naïve Bayes' probability of playing if it's **sunny**?

$$\begin{aligned} p(c \mid x) &= p(c)p(x \mid c)/p(x) \\ p(\text{yes} \mid \text{sunny}) &= p(\text{yes})p(\text{sunny} \mid \text{yes})/p(sunny) \end{aligned}$$

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14 [2]	5/14

[1] $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

[2]
$$p(c) = p(\text{yes}) = 9/14 = 0.64$$
$$p(x) = p(\text{sunny}) = 5/14 = 0.36$$

What is the Naïve Bayes' probability of playing if it's **sunny**?

$$\begin{aligned} p(c \mid x) &= p(c)p(x \mid c)/p(x) \\ p(\text{yes} \mid \text{sunny}) &= p(\text{yes})p(\text{sunny} \mid \text{yes})/p(sunny) \\ p(\text{yes} \mid \text{sunny}) &= 0.64 * 0.33/0.36 \end{aligned}$$

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

Likelihood table

	Play	
Outlook	**yes**	**no**
sunny	3/9 [1]	2/5
overcast	4/9	0/5
rainy	2/9	3/5
	9/14 [2]	5/14

[1] $p(x \mid c) = p(\text{sunny} \mid \text{yes}) = 3/9 = 0.33$

[2] $$p(c) = p(\text{yes}) = 9/14 = 0.64$$
$$p(x) = p(\text{sunny}) = 5/14 = 0.36$$

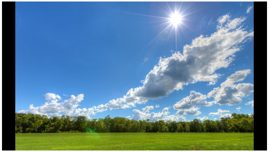
What is the Naïve Bayes' probability of playing if it's **sunny**?

$$\begin{aligned} p(c \mid x) &= p(c)p(x \mid c)/p(x) \\ p(\text{yes} \mid \text{sunny}) &= p(\text{yes})p(\text{sunny} \mid \text{yes})/p(sunny) \\ p(\text{yes} \mid \text{sunny}) &= 0.64 * 0.33/0.36 \\ p(\text{yes} \mid \text{sunny}) &= 0.59 \end{aligned}$$

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I play *calcio* today?

If. . .  **bring the ball!**

Naïve Bayes

A toy example: Should I play *calcio* today?

Considering more data

Outlook	**Temp**	**Humidity**	**Windy**	**Play**
rainy	hot	high	false	no
rainy	hot	high	true	no
overcast	hot	high	false	yes
sunny	mild	high	false	yes
sunny	cool	normal	false	yes
sunny	cool	normal	true	no
overcast	cool	normal	true	yes
rainy	mild	high	false	no
rainy	cool	normal	false	yes
sunny	mild	normal	false	yes
rainy	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
sunny	mild	high	true	no

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I **play** *calcio* today?

Frequency tables

Outlook	yes	no
sunny	3	2
overcast	4	0
rainy	2	3

Likelihood tables

Outlook	yes	no
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I **play** *calcio* today?

Frequency tables

Outlook	**yes**	**no**
sunny	3	2
overcast	4	0
rainy	2	3

Humid	**yes**	**no**
high	3	4
normal	6	1

Likelihood tables

Outlook	**yes**	**no**
sunny	$3/9$	$2/5$
overcast	$4/9$	$0/5$
rainy	$2/9$	$3/5$

Humid	**yes**	**no**
high	$3/9$	$4/5$
normal	$6/9$	$1/5$

Adapted from http://www.saedsayad.com/naive_bayesian.htm

Naïve Bayes

A toy example: Should I **play** *calcio* today?

Frequency tables

Outlook	**yes**	**no**
sunny	3	2
overcast	4	0
rainy	2	3

Humid	**yes**	**no**
high	3	4
normal	6	1

Temp	**yes**	**no**
hot	2	2
mild	4	2
cool	3	1

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

A toy example: Should I **play** *calcio* today?

Frequency tables

Outlook	**yes**	**no**
sunny	3	2
overcast	4	0
rainy	2	3

Humid	**yes**	**no**
high	3	4
normal	6	1

Temp	**yes**	**no**
hot	2	2
mild	4	2
cool	3	1

Windy	**yes**	**no**
false	6	2
true	3	3

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

Adapted from `http://www.saedsayad.com/naive_bayesian.htm`

Naïve Bayes

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

Naïve Bayes

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

outlook	**temp**	**humidity**	**windy**	play
rainy	cool	high	true	?

Naïve Bayes

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

outlook	**temp**	**humidity**	**windy**	play
rainy	cool	high	true	?

$$p(\text{yes} \mid x) = \frac{p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes})}{p(\text{rainy})p(\text{cool})p(\text{high})p(\text{true})}$$

Naïve Bayes

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

outlook	**temp**	**humidity**	**windy**	play
rainy	cool	high	true	?

$$
\begin{aligned}
p(\text{yes} \mid x) &= \frac{p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes})}{p(\text{rainy})p(\text{cool})p(\text{high})p(\text{true})} \\
&= \frac{9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9}{5/14 \times 4/14 \times 7/14 \times 6/14}
\end{aligned}
$$

Naïve Bayes

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

outlook	**temp**	**humidity**	**windy**	play
rainy	cool	high	true	?

$$
\begin{aligned}
p(\text{yes} \mid x) &= \frac{p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes})}{p(\text{rainy})p(\text{cool})p(\text{high})p(\text{true})} \\
&= \frac{9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9}{5/14 \times 4/14 \times 7/14 \times 6/14} \\
&= 0.00529/0.02811
\end{aligned}
$$

Naïve Bayes

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

outlook	**temp**	**humidity**	**windy**	play
rainy	cool	high	true	?

$$
\begin{aligned}
p(\text{yes} \mid x) &= \frac{p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes})}{p(\text{rainy})p(\text{cool})p(\text{high})p(\text{true})} \\
&= \frac{9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9}{5/14 \times 4/14 \times 7/14 \times 6/14} \\
&= 0.00529/0.02811 = 0.188 \sim 0.2
\end{aligned}
$$

Naïve Bayes

Likelihood tables

Outlook	**yes**	**no**
sunny	3/9	2/5
overcast	4/9	0/5
rainy	2/9	3/5

Temp	**yes**	**no**
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humid	**yes**	**no**
high	3/9	4/5
normal	6/9	1/5

Windy	**yes**	**no**
false	6/9	2/5
true	3/9	3/5

outlook	**temp**	**humidity**	**windy**	play
rainy	cool	high	true	?

$$
\begin{aligned}
p(\text{yes} \mid x) &= \frac{p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes})}{p(\text{rainy})p(\text{cool})p(\text{high})p(\text{true})} \\
&= \frac{9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9}{5/14 \times 4/14 \times 7/14 \times 6/14} \\
&= 0.00529/0.02811 = 0.188 \sim 0.2 \textbf{ no game}
\end{aligned}
$$

Naïve Bayes

Back to the math. . .

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{8}$$

Naïve Bayes

Back to the math...

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{8}$$

The probability $p(\mathbf{x})$ is constant for any given input!

Naïve Bayes

Back to the math...

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{8}$$

The probability $p(\mathbf{x})$ is constant for any given input!

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{\cancel{p(\mathbf{x})}} \tag{9}$$

Naïve Bayes

Back to the math...

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{p(\mathbf{x})} \tag{8}$$

The probability $p(\mathbf{x})$ is constant for any given input!

$$p(C \mid \mathbf{x}) = \frac{p(C)\ p(\mathbf{x} \mid C)}{\cancel{p(\mathbf{x})}} \tag{9}$$

$$p(c \mid \mathbf{x}) \propto p(c)p(\mathbf{x} \mid c) \tag{10}$$

Naïve Bayes

Back to the math...

$$p(c \mid \mathbf{x}) \propto p(c)p(\mathbf{x} \mid c) \tag{11}$$

Naïve Bayes

Back to the math...

$$p(c \mid \mathbf{x}) \propto p(c)p(\mathbf{x} \mid c) \tag{11}$$

But $\mathbf{x}$ is a vector

Naïve Bayes

Back to the math...

$$p(c \mid \mathbf{x}) \propto p(c)p(\mathbf{x} \mid c) \tag{11}$$

But **x** is a vector

$$p(c \mid x_1 \ldots x_n) \propto p(c)p(x_1 \mid c) \times p(x_2 \mid c) \times \ldots \times p(x_n \mid c) \tag{12}$$

Naïve Bayes

Back to the math. . .

$$p(c \mid \mathbf{x}) \propto p(c)p(\mathbf{x} \mid c) \tag{11}$$

But **x** is a vector

$$p(c \mid x_1 \dots x_n) \propto p(c)p(x_1 \mid c) \times p(x_2 \mid c) \times \dots \times p(x_n \mid c) \tag{12}$$

Eq.(12) can be rewritten as

$$p(c \mid x_1 \dots x_n) \propto p(c) \prod_{i=1}^{n} p(x_i \mid c) \tag{13}$$

Naïve Bayes

The classification process

Back to the toy example

$$p(\text{yes} \mid x) \quad \propto \quad p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes})$$

Back to the toy example

$$
\begin{aligned}
p(\text{yes} \mid x) &\propto p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes}) \\
&\propto 9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9
\end{aligned}
$$

Naïve Bayes

The classification process

Back to the toy example

$$
\begin{aligned}
p(\text{yes} \mid x) &\propto p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes}) \\
&\propto 9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9 \\
&\propto 0.00529
\end{aligned}
$$

Naïve Bayes

The classification process

Back to the toy example

$$
\begin{aligned}
p(\text{yes} \mid x) &\propto p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes}) \\
&\propto 9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9 \\
&\propto 0.00529 \textbf{ not a probability!}
\end{aligned}
$$

Back to the toy example

$$
\begin{aligned}
p(\text{yes} \mid x) &\propto p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes}) \\
&\propto 9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9 \\
&\propto 0.00529 \ \textbf{not a probability!}
\end{aligned}
$$

Classification: the maximum for all the classes

$$
c \propto \arg\max_{c} p(c) \prod_{i=1}^{n} p(x_i \mid c) \tag{14}
$$

Naïve Bayes

The classification process

Back to the toy example

$$\begin{aligned} p(\text{yes} \mid x) &\propto p(\text{yes})p(\text{rainy} \mid \text{yes})p(\text{cool} \mid \text{yes})p(\text{high} \mid \text{yes})p(\text{true} \mid \text{yes}) \\ &\propto 9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9 \\ &\propto 0.00529 \textbf{ not a probability!} \end{aligned}$$

Classification: the maximum for all the classes

$$c \propto \arg\max_c p(c) \prod_{i=1}^{n} p(x_i \mid c) \tag{14}$$

```
compute p(yes|x)
compute p(no|x)
if p(yes|x) > p(no|x):
    yes
else:
    no
```

Training a Machine Learning Model

The dataset

We need a bunch of documents with their associated **class**

The dataset

We need a bunch of documents with their associated **class**

kind	**examples**

The dataset

We need a bunch of documents with their associated **class**

kind	**examples**
binary	{positive, negative}
	{0, 1}
	{-1, 1}

The dataset

We need a bunch of documents with their associated **class**

kind	examples
binary	{positive, negative}
	{0, 1}
	{-1, 1}
multiclass	{positive, neutral, negative}
	{0,1,2}

The dataset

We need a bunch of documents with their associated **class**

kind	**examples**
binary	{positive, negative}
	{0, 1}
	{-1, 1}
multiclass	{positive, neutral, negative}
	{0,1,2}

In our case, we need the positivity sentiment:

d_1	pos	d_5	neg	d_9	neu
d_2	neu	d_6	neg	d_{10}	pos
d_3	pos	d_7	neg	d_{11}	neu
d_4	pos	d_8	pos	d_{12}	neg

The dataset

Option 1 You use a corpus created by somebody else

The dataset

Option 1 You use a corpus created by somebody else

Option 2 You build your own corpus

The dataset

Option 1 You use a corpus created by somebody else

Option 2 You build your own corpus

(a) You have/hire an expert to do it

The dataset

Option 1 You use a corpus created by somebody else

Option 2 You build your own corpus

(a) You have/hire an expert to do it
(b) You engage non-experts through gamification

The dataset

Option 1 You use a corpus created by somebody else

Option 2 You build your own corpus

(a) You have/hire an expert to do it
(b) You engage non-experts through gamification
(c) You hire non-experts through explicit crowdsourcing

The dataset

Option 1 You use a corpus created by somebody else

Option 2 You build your own corpus

a. You have/hire an expert to do it
b. You engage non-experts through gamification
c. You hire non-experts through explicit crowdsourcing
d. There are many other ways to get annotated data

The dataset

Option 1 **You use a corpus created by somebody else**

Option 2 You build your own corpus

a. You have/hire an expert to do it
b. You engage non-experts through gamification
c. You hire non-experts through explicit crowdsourcing
d. There are many other ways to get annotated data

Let us go and build a classifier with a corpus built by Hutto and Gilbert (2014)[2]

[2] http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

Let us go and build a classifier with a corpus built by Hutto and Gilbert (2014)[2]

For this, you have to download and install the software companion of NLP in Action:

`https://github.com/totalgood/nlpia`

[2] http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

What I did on OsX

I use pipenv[3]

```
$ pipenv install --skip-lock nlpia
```

On Github they explain how to install it with conda or pip if you plan to contribute to the project

[3] https://pipenv.readthedocs.io/en/latest/

References

Hutto, C. and E. Gilbert
2014. VADER:A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, Ann Arbor, MI.

Lane, H., C. Howard, and H. Hapkem
2019. *Natural Language Processing in Action*. Shelter Island, NY: Manning Publication Co.

Maron, M.
1961. Automatic indexing: An experimental inquiry. *Journal of the ACM*, 8:404–417.