Naïve Bayes

Daniel Nogueira

dnogueira@ipca.pt

Naïve Bayes

Spam Classification

Given an email, predict whether it is spam or not

Medical Diagnosis

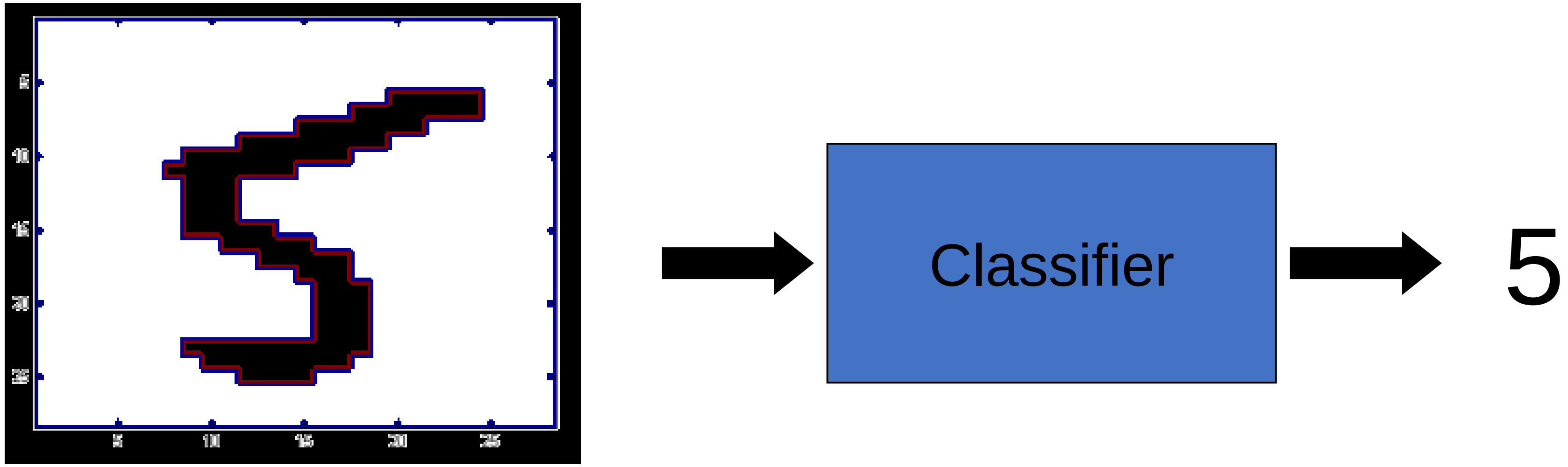
Given a list of symptoms, predict whether a patient has disease X or not

Weather

Based on temperature, humidity, etc… predict if it will rain tomorrow

Naïve Bayes

Digit Recognition



- $X_1, \ldots, X_n \in \{0,1\}$ **(Black vs. White pixels)**

Naïve Bayes

There are three methods to establish a classifier

a) Model a classification rule directly

Examples: k-NN, decision trees, perceptron, SVM

Naïve Bayes

There are three methods to establish a classifier

a) Model a classification rule directly

Examples: k-NN, decision trees, perceptron, SVM

b) Model the probability of class memberships given input data

Example: multi-layered perceptron with the cross-entropy cost

Naïve Bayes

There are three methods to establish a classifier

a) Model a classification rule directly

Examples: k-NN, decision trees, perceptron, SVM

b) Model the probability of class memberships given input data

Example: multi-layered perceptron with the cross-entropy cost

c) Make a probabilistic model of data within each class

Examples: naive Bayes, model based classifiers

Naïve Bayes

There are three methods to establish a classifier

a) Model a classification rule directly *discriminative classification*

Examples: k-NN, decision trees, perceptron, SVM

b) Model the probability of class memberships given input data *probabilistic classification*

Example: multi-layered perceptron with the cross-entropy cost

c) Make a probabilistic model of data within each class *probabilistic classification*

Examples: naive Bayes, model based classifiers

Naïve Bayes

Teorema de Naive Bayes

- 100 pessoas realizaram o teste.
- 20% das pessoas que realizaram o teste possuíam a doença.
- 90% das pessoas que possuíam a doença, receberam positivo no teste.
- 30% das pessoas que não possuíam a doença, receberam positivo no teste.

Naïve Bayes

Teorema de Naive Bayes

- 100 pessoas realizaram o teste.
- 20% das pessoas que realizaram o teste possuíam a doença.
- 90% das pessoas que possuíam a doença, receberam positivo no teste.
- 30% das pessoas que não possuíam a doença, receberam positivo no teste.

Se uma nova pessoa realizar o teste e receber um resultado positivo, qual a probabilidade de ela possuir a doença?

Naïve Bayes

Teorema de Naive Bayes

- 100 pessoas realizaram o teste.
- 20% das pessoas que realizaram o teste possuíam a doença.
- 90% das pessoas que possuíam a doença, receberam positivo no teste.
- 30% das pessoas que não possuíam a doença, receberam positivo no teste.

P(doença|positivo) = 20% * 90%
P(doença|positivo) = 0,2 * 0,9
P(doença|positivo) = 0,18

Se uma nova pessoa realizar o teste e receber um resultado positivo, qual a probabilidade de ela possuir a doença?

Naïve Bayes

Teorema de Naive Bayes

- 100 pessoas realizaram o teste.
- 20% das pessoas que realizaram o teste possuíam a doença.
- 90% das pessoas que possuíam a doença, receberam positivo no teste.
- 30% das pessoas que não possuíam a doença, receberam positivo no teste.

Se uma nova pessoa realizar o teste e receber um resultado positivo, qual a probabilidade de ela possuir a doença?

P(doença|positivo) = 20% * 90%
P(doença|positivo) = 0,2 * 0,9
P(doença|positivo) = 0,18

P(não doença|positivo) = 80% * 30%
P(não doença|positivo) = 0,8 * 0,3
P(não doença|positivo) = 0,24

Naïve Bayes

Teorema de Naive Bayes

Se uma nova pessoa realizar o teste e receber um resultado positivo, qual a probabilidade de ela possuir a doença?

- 100 pessoas realizaram o teste.
- 20% das pessoas que realizaram o teste possuíam a doença.
- 90% das pessoas que possuíam a doença, receberam positivo no teste.
- 30% das pessoas que não possuíam a doença, receberam positivo no teste.

P(doença|positivo) = 20% * 90%
P(doença|positivo) = 0,2 * 0,9
P(doença|positivo) = 0,18

P(não doença|positivo) = 80% * 30%
P(não doença|positivo) = 0,8 * 0,3
P(não doença|positivo) = 0,24

P(doença|positivo) = 0,18/(0,18+0,24) = 0,4285

P(não doença|positivo) = 0,24/(0,18+0,24) = 0,5714

0,4285 + 0,5714 = 0,9999.. ou aproximadamente 1.

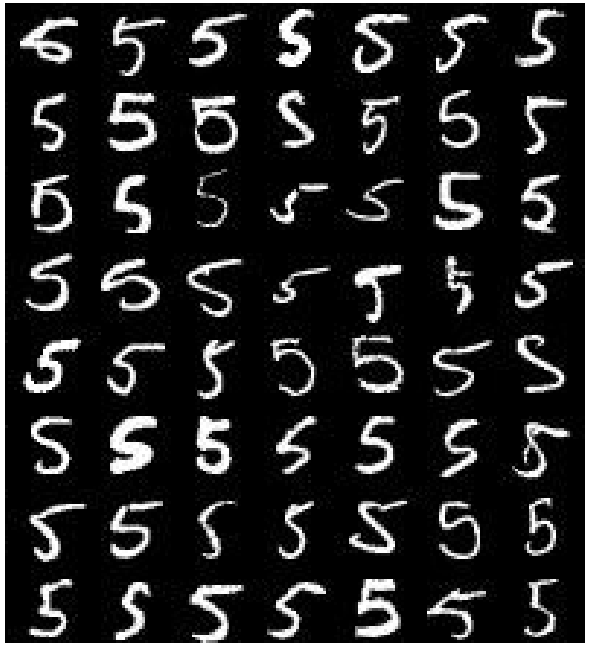
Naïve Bayes

Teorema de Naive Bayes

$$P(Y|X_1, \ldots, X_n) = \frac{P(X_1, \ldots, X_n|Y)P(Y)}{P(X_1, \ldots, X_n)}$$

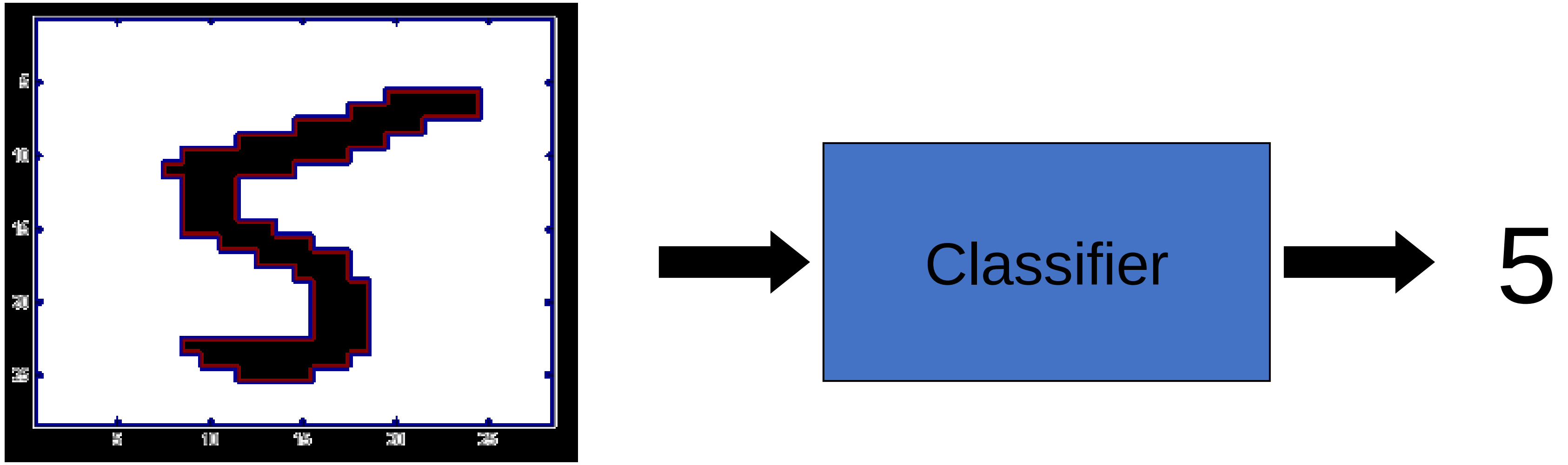
Naïve Bayes

Digit Recognition



Naïve Bayes

Digit Recognition



- $X_1,\ldots,X_n \in \{0,1\}$ (Black vs. White pixels)
- $Y \in \{5,6\}$ (predict whether a digit is a 5 or a 6)

Naïve Bayes

$X_1,\ldots,X_n \in \{0,1\}$ (Black vs. White pixels)

$$
\begin{aligned}
P(Y=5|X_1,\ldots,X_n) &= \frac{P(X_1,\ldots,X_n|Y=5)P(Y=5)}{P(X_1,\ldots,X_n|Y=5)P(Y=5)+P(X_1,\ldots,X_n|Y=6)P(Y=6)}\\
P(Y=6|X_1,\ldots,X_n) &= \frac{P(X_1,\ldots,X_n|Y=6)P(Y=6)}{P(X_1,\ldots,X_n|Y=5)P(Y=5)+P(X_1,\ldots,X_n|Y=6)P(Y=6)}
\end{aligned}
$$

Naïve Bayes

$X_1,\dots,X_n \in \{0,1\}$ (Black vs. White pixels)

$$
\begin{aligned}
P(Y=5|X_1,\dots,X_n) &= \frac{P(X_1,\dots,X_n|Y=5)P(Y=5)}{P(X_1,\dots,X_n|Y=5)P(Y=5)+P(X_1,\dots,X_n|Y=6)P(Y=6)} \\
P(Y=6|X_1,\dots,X_n) &= \frac{P(X_1,\dots,X_n|Y=6)P(Y=6)}{P(X_1,\dots,X_n|Y=5)P(Y=5)+P(X_1,\dots,X_n|Y=6)P(Y=6)}
\end{aligned}
$$

To classify, we'll simply compute these two probabilities and predict based on which one is greater !!!!!

Naïve Bayes

Play Tennis

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Naïve Bayes

Play Tennis

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Outlook	Play=*Yes*	Play=*No*
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Play=*Yes – 9*

Play=*No – 5*

Naïve Bayes

Play Tennis

Outlook	Play=*Yes*	Play=*No*
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	Play=*Yes*	Play=*No*
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Humidity	Play=*Yes*	Play=*No*
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=*Yes*	Play=*No*
Strong	3/9	3/5
Weak	6/9	2/5

$P(\text{Play}=Yes) = 9/14 \qquad P(\text{Play}=No) = 5/14$

Naïve Bayes

Play Tennis

Given a new instance,

$\mathbf{x}'$=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)

Naïve Bayes

Play Tennis

Given a new instance,

$\mathbf{x}'$=(Outlook=*Sunny,* Temperature=*Cool,* Humidity=*High,* Wind=*Strong*)

– Look up tables

P(Outlook=*Sunny*|Play=*Yes*) = 2/9

P(Temperature=*Cool*|Play=*Yes*) = 3/9

P(Huminity=*High*|Play=*Yes*) = 3/9

P(Wind=*Strong*|Play=*Yes*) = 3/9

P(Play=*Yes*) = 9/14

P(Outlook=*Sunny*|Play=*No*) = 3/5

P(Temperature=*Cool*|Play==*No*) = 1/5

P(Huminity=*High*|Play=*No*) = 4/5

P(Wind=*Strong*|Play=*No*) = 3/5

P(Play=*No*) = 5/14

Naïve Bayes

Play Tennis

Given a new instance,

$\mathbf{x}'$=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)

- Look up tables

P(Outlook=*Sunny*|Play=*Yes*) = 2/9

P(Temperature=*Cool*|Play=*Yes*) = 3/9

P(Huminity=*High*|Play=*Yes*) = 3/9

P(Wind=*Strong*|Play=*Yes*) = 3/9

P(Play=*Yes*) = 9/14

P(Outlook=*Sunny*|Play=*No*) = 3/5

P(Temperature=*Cool*|Play==*No*) = 1/5

P(Huminity=*High*|Play=*No*) = 4/5

P(Wind=*Strong*|Play=*No*) = 3/5

P(Play=*No*) = 5/14

- MAP rule

P(*Yes*|$\mathbf{x}'$): [P(*Sunny*|Yes)P(*Cool*|*Yes*)P(*High*|Yes)P(*Strong*|*Yes*)]P(Play=*Yes*) = 0.0053

P(*No*|$\mathbf{x}'$): [P(*Sunny*|No) P(*Cool*|No)P(*High*|*No*)P(*Strong*|*No*)]P(Play=*No*) = 0.0206

Naïve Bayes

Play Tennis

Given a new instance,

$\mathbf{x}'$=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)

- Look up tables

P(Outlook=*Sunny*|Play=*Yes*) = 2/9

P(Temperature=*Cool*|Play=*Yes*) = 3/9

P(Huminity=*High*|Play=*Yes*) = 3/9

P(Wind=*Strong*|Play=*Yes*) = 3/9

P(Play=*Yes*) = 9/14

P(Outlook=*Sunny*|Play=*No*) = 3/5

P(Temperature=*Cool*|Play==*No*) = 1/5

P(Huminity=*High*|Play=*No*) = 4/5

P(Wind=*Strong*|Play=*No*) = 3/5

P(Play=*No*) = 5/14

- MAP rule

P(*Yes*|$\mathbf{x}'$): [P(*Sunny*|Yes)P(*Cool*|*Yes*)P(*High*|Yes)P(*Strong*|*Yes*)]P(Play=*Yes*) = 0.0053

P(*No*|$\mathbf{x}'$): [P(*Sunny*|No) P(*Cool*|No)P(*High*|*No*)P(*Strong*|*No*)]P(Play=*No*) = 0.0206

Given the fact P(*Yes*|$\mathbf{x}'$) < P(*No*|$\mathbf{x}'$), we label $\mathbf{x}'$ to be "*No*".

Naïve Bayes

Advantages

- It has a simple mathematical equation and is therefore a “fast” algorithm.
- It is efficient in multi-class predictions.
- In problems where the assumption of independence between attributes is valid, it can achieve better performance than other more complex methods.
- It performs well when subjected to categorical variables.

Disadvantages

- In real life, it is almost impossible to have a set of indicators that are completely independent (correlation between variables exists).
- It is considered a poor estimator, since the probabilities calculated by the algorithm should not be considered accurate.
- If the categorical variable has a category (in the test dataset) that was not observed in the training dataset, then the model will assign a probability of 0 (zero) and will not be able to make a prediction.
- This is often known as “zero frequency”. To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is the so-called Laplacian correction.

Naïve Bayes

Conclusions

- Naïve Bayes based on the independence assumption

Training is very easy and fast; just requiring considering each attribute in each class separately

Test is straightforward; just looking up tables or calculating conditional probabilities with normal distributions

Naïve Bayes

Conclusions

- Naïve Bayes based on the independence assumption

Training is very easy and fast; just requiring considering each attribute in each class separately

Test is straightforward; just looking up tables or calculating conditional probabilities with normal distributions

- A popular model

Performance competitive to most of state-of-the-art classifiers even in presence of violating independence assumption

Many successful applications, e.g., spam mail filtering.

Naïve Bayes

Conclusions

- Naïve Bayes based on the independence assumption

Training is very easy and fast; just requiring considering each attribute in each class separately

Test is straightforward; just looking up tables or calculating conditional probabilities with normal distributions

- A popular model

Performance competitive to most of state-of-the-art classifiers even in presence of violating independence assumption

Many successful applications, e.g., spam mail filtering.

Issues

- Violation of Independence Assumption

Naïve Bayes

Conclusions

- Naïve Bayes based on the independence assumption

Training is very easy and fast; just requiring considering each attribute in each class separately

Test is straightforward; just looking up tables or calculating conditional probabilities with normal distributions

- A popular model

Performance competitive to most of state-of-the-art classifiers even in presence of violating independence assumption

Many successful applications, e.g., spam mail filtering.

Issues

- Violation of Independence Assumption
- Zero conditional probability Problem

Daniel Nogueira

dnogueira@ipca.pt