DM I: Block 'Classification'

Unit 'Naive Bayes'

Myra Spiliopoulou

FACULTY OF
COMPUTER SCIENCE

EXAMPLE: 'patient responses' dataset

Simplified version:

Id	I2	I15	I30	I22	I31	I9	I26	Response
#1	f	VH	yes	better	no	r	no	yes
#2	m	M	no	better	no	b	yes	no
#3	m	M	no	worse	no	b	no	no
#4	f	VH	yes	worse	no	b	no	yes
#5	m	L	no	no effect	no	l	no	no
#6	m	M	no	better	no	l	no	no
#7	f	VH	yes	better	yes	l	yes	yes
#8	f	H	no	better	yes	r	no	no
#9	f	H	yes	better	no	l	no	yes
#10	m	M	yes	worse	no	b	no	no
#11	m	M	no	no effect	no	l	no	no
#12	f	H	no	no effect	no	r	no	no
#13	f	L	yes	better	no	l	yes	yes
#14	m	M	no	worse	no	b	no	no
#15	m	L	no	no effect	no	l	yes	no

EXAMPLE: 'patient responses' dataset

and version with both categorical and numerical attributes:

Id	I2	I15	I30	I22	I31	I9	I26	Response
#1	f	VH	yes	1	no	r	no	yes
#2	m	M	no	2	no	b	yes	no
#3	m	M	no	6	no	b	no	no
#4	f	VH	yes	7	no	b	no	yes
#5	m	L	no	3	no	l	no	no
#6	m	M	no	1	no	l	no	no
#7	f	VH	yes	1	yes	l	yes	yes
#8	f	H	no	2	yes	r	no	no
#9	f	H	yes	2	no	l	no	yes
#10	m	M	yes	6	no	b	no	no
#11	m	M	no	3	no	l	no	no
#12	f	H	no	4	no	r	no	no
#13	f	L	yes	4	no	l	yes	yes
#14	m	M	no	7	no	b	no	no
#15	m	L	no	5	no	l	yes	no

Classification with Naive Bayes from Witten & Eibe

Law of Bayes

The probability of an event H given evidence E is: $p(H|E) = \frac{p(E|H)p(H)}{p(E)}$

Classification with Naive Bayes from Witten & Eibe

Law of Bayes

The probability of an event H given evidence E is: $p(H|E) = \frac{p(E|H)p(H)}{p(E)}$

Law of Bayes, using the Independence Assumption

The probability of an event H given evidence E is:

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)} = \frac{\prod_{i=1}^{n} p(E_i|H)p(H)}{p(E)}$$

i.e. it is assumed that $p(E|H) = \prod_{i=1}^{n} p(E_i|H)$ (naive).

Appeared in:

"Essay towards solving a problem in the doctrine of chances" (1763)

by Thomas Bayes (born: 1702, London; died: 1761, Tunbridge Wells, Kent)

Classification with Naive Bayes

Witten & Eibe

Law of Bayes, using the Independence Assumption

The probability of an event H given evidence E is:

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)} = \frac{\prod_{i=1}^{n} p(E_i|H)p(H)}{p(E)}$$

i.e. it is assumed that $p(E|H) = \prod_{i=1}^{n} p(E_i|H)$ (naive).

Classification with Naive Bayes

Witten & Eibe

Law of Bayes, using the Independence Assumption

The probability of an event H given evidence E is:

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)} = \frac{\prod_{i=1}^{n} p(E_i|H)p(H)}{p(E)}$$

i.e. it is assumed that $p(E|H) = \prod_{i=1}^{n} p(E_i|H)$ (naive).

In classification, an event is the observation of a class H, while E is a record, composed of attributes values $E_1, \ldots, E_n$ in an n-dimensional feature space $A = \{a_1, \ldots, a_n\}$.

Classification with Naive Bayes

Witten & Eibe

Law of Bayes, using the Independence Assumption

The probability of an event H given evidence E is:

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)} = \frac{\prod_{i=1}^{n} p(E_i|H)p(H)}{p(E)}$$

i.e. it is assumed that $p(E|H) = \prod_{i=1}^{n} p(E_i|H)$ (naive).

In classification, an event is the observation of a class H,
while E is a record, composed of attributes values $E_1, \ldots, E_n$ in an n-dimensional feature space $A = \{a_1, \ldots, a_n\}$.
The independence assumption means that:

- All attributes contribute equally to the class prediction.
- Within a given class, the value of an attribute does not influence the values of other attributes.

Classification with Naive Bayes

Learning phase of Naive Bayes

For each (attribute, value)-pair (a, z) for an attribute $a \in A$:

- For each label $y \in L$: compute $p((a, z)|y)$

Classification with Naive Bayes

Learning phase of Naive Bayes

For each (attribute, value)-pair (a, z) for an attribute $a \in A$:

- For each label $y \in L$: compute $p((a, z)|y)$

Application phase with Naive Bayes

For an instance x with unknown label:

- For each label $y \in L$: compute $p(y|x)$ using the computations of the learning phase.
- Assign to x the label with the highest probability, i.e. $label(x) = \arg\max_{y \in L} p(y|x)$.

EXAMPLE for the 'patient responses' dataset (categorical attributes only)

Priors of the target variable: $p(yes) = 5/15$, $p(no) = 10/15$.

1. Count the occurrences of each class given the value of each attribute:

I2:	*yes*	*no*
f	5	2
m	0	8

I15:	*yes*	*no*
VH	3	0
H	1	2
M	0	6
L	1	2

I30:	*yes*	*no*
yes	5	1
no	0	9

I22:	*yes*	*no*
better	4	3
worse	1	3
no effect	0	4

I31:	*yes*	*no*
yes	1	1
no	4	9

I9:	*yes*	*no*
r	1	2
l	3	4
b	1	4

I26:	*yes*	*no*
yes	2	2
no	3	8

2. Compute the NB model

3.Apply the NB model

To which class should we assign $x =<$#85,f,VH, yes, no effect, no, l, no$>$?

Classification with Naive Bayes – The Zero-Frequency Problem

For a training set D from a population $\mathscr{D}$, a Labelset $L = \{y_1, \ldots, y_k\}$, an attribute $a \in A$ and a $y \in L$, let

$$D_y = \{x \in D | label(x) = y\}, D_{(a,z_i)} = \{x \in D | x.a = z_i\}$$

Classification with Naive Bayes – The Zero-Frequency Problem

For a training set D from a population $\mathscr{D}$, a Labelset $L = \{y_1, \ldots, y_k\}$, an attribute $a \in A$ and a $y \in L$, let

$$D_y = \{x \in D | label(x) = y\}, D_{(a,z_i)} = \{x \in D | x.a = z_i\}$$

If there is a value z_i that a can take and a label $y \in L$ so that $p((a, z_i)|y) = 0$, then
for all $x \in \mathscr{D}$ with $x.a = z_i$ it holds that: $p(y|x) = 0$.

Classification with Naive Bayes – The Zero-Frequency Problem

For a training set D from a population $\mathscr{D}$, a Labelset $L = \{y_1, \ldots, y_k\}$, an attribute $a \in A$ and a $y \in L$, let

$$D_y = \{x \in D | label(x) = y\}, D_{(a,z_i)} = \{x \in D | x.a = z_i\}$$

If there is a value z_i that a can take and a label $y \in L$ so that $p((a,z_i)|y) = 0$, then
for all $x \in \mathscr{D}$ with $x.a = z_i$ it holds that: $p(y|x) = 0$.

Laplace Estimator

Instead of computing $p((a,z_i)|y)$ as $\frac{|D_{(a,z_i)} \cap D_y|}{|D_y|}$, we set:

$$p((a,z_i)|y) = \frac{|D_{(a,z_i)} \cap D_y| + \frac{w}{n_a}}{|D_y| + w}$$

where w is some small weight > 0 and n_a is the number of distinct values that attribute a can take.
This is a generalization of the original Laplace estimator, where $w = n_a$.

Classification with Naive Bayes – more on missing values

Let $x \in \mathscr{D}$ be an instance of unknown label, so that the value of x for some attribute a is missing.

How do we deduce the most likely class of x?

Naive Bayes for attributes with continuous valueranges

For a training set D from a population $\mathscr{D}$, a Labelset $L = \{y_1, \ldots, y_k\}$, let $a \in A$ be an attribute with a continuous valuerange, and let $D_a \subseteq D$ be the set of instances from D, that have a non-NULL value for attribute a.

Using densities in the NB calculations – from Witten & Eibe

Assumption: The values of a follow a Gaussian distribution [a] (also: normal distribution), with density function:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{(x-\mu_a)^2}{2\sigma_a^2}}$$

where μ_a is the *sample mean*: $\mu_a = \frac{1}{|D_a|} \sum_{x \in D_a} x.a$
σ_a is the *sample standard deviation*, and
σ_a^2 the *sample variance*: $\sigma_a^2 = \frac{1}{|D_a|-1} \sum_{x \in D_a} (x.a - \mu_a)^2$

[a] Named after the German mathematician Karl Friedrich Gauss (1777-1855), who laid the foundations of number theory.

EXAMPLE for the 'patient responses' dataset (mixed attributes)

Priors of the target variable are $p(yes) = 5/15$, $p(no) = 10/15$; the computations for all attributes except I22 remain the same.

I22:	*yes*	*no*
	1, 2, 4, 7	1, 2, 3, 4, 5, 6, 7
	$\mu = ?, \sigma = ?$	$\mu = ?, \sigma = ?$

To which class should we assign $x =$<#86,f,VH, yes, 4, no, l, no> ?

Progress and outlook

We have seen:

- $\surd$ How to train a classification model with help of the Law of Bayes
- $\surd$ How to deal with values that may occur in the test set and did not occur in the training set
- $\surd$ How to extend the algorithm for learning on mixed (categorical and numerical) attributes

NB classifiers perform usually well, despite the naive assumption, unless

- $\rightarrow$ there are many redundant attributes, or
- $\rightarrow$ the valueranges do not follow a Gaussian distribution

$\Downarrow$

- How to figure out how good a classifier is?

Thank you very much!

Questions?