

19T3 : COMP9417 Machine Learning and Data Mining

Lectures: Classification (2)

Topic: Questions from lecture

Version: with answers

Last revision: 10 October 2019

Introduction

A question and an exercise from the course lecture on the Naive Bayes classifier. Do these then complete the accompanying notebook.

Naive Bayes classifier

Question 1 Consider the example application of Bayes Theorem on slides 21-24 in the lecture notes.

Now suppose the a second laboratory test is ordered for the same patient, and this test also returns a positive result. What are the posterior probabilities of *cancer* and $\neg\text{cancer}$ following these two tests? Note: you can assume that the two tests are independent.

Answer

Assume: two tests are independent.

$$\begin{array}{ll}
 \text{Priors} & P(\text{cancer}) = 0.008 \quad P(\text{not cancer}) = 0.992 \\
 \text{Likelihoods} & P(\oplus | \text{cancer}) = 0.98 \quad P(\ominus | \text{cancer}) = 0.02 \quad \text{Given} \\
 & P(\oplus | \text{not cancer}) = 0.03 \quad P(\ominus | \text{not cancer}) = 0.97
 \end{array}$$

First test = \oplus $\left[\arg \max_h P(h | D) = P(D|h)P(h) \right]$

$$\begin{aligned}
 P(\oplus | \text{cancer}) P(\text{cancer}) &= 0.98 \times 0.008 = 0.0078 \\
 P(\oplus | \text{not cancer}) P(\text{not cancer}) &= 0.03 \times 0.992 = 0.0298
 \end{aligned}$$

Second test = \oplus $D = \{\oplus, \ominus\}$ $\Rightarrow \text{MAP} = \text{not cancer}$

$$P(d_1 | h) P(d_2 | h) P(h) \quad // \text{By independence of data!}$$

$$P(\oplus | c) P(\oplus | \text{not } c) P(c) = 0.98 \times 0.98 \times 0.008 = 0.0077$$

$$P(\oplus | \text{not } c) P(\oplus | c) P(\text{not } c) = 0.03 \times 0.03 \times 0.992 = 0.0009$$

$$\Rightarrow \text{MAP} = \text{cancer}$$

N.B. how quickly probabilities get v. small: \Rightarrow log-space calculation

Question 2 Work through the example of applying Naive Bayes to text on slides 71-77. Be sure you are clear on the difference between the multinomial and multivariate Bernoulli models.

Answer

See notes on following pages. First up are two pages with the key steps. The rest is some older handwritten notes with some extra information, but at the cost of poorer legibility – these are just included for completeness.

Question 6 : \oplus

$e_1 : b \ d \ e \ b \ b \ d \ e$

$e_2 : b \ c \ e \ b \ b \ d \ d \ e \ c \ c$

$e_3 : a \ d \ a \ d \ e \ a \ e \ e$

$e_4 : b \ a \ d \ b \ e \ d \ a \ b$

MULTINOMIAL
Count Vector

	a	b	c	class
e_1	0	3	0	+
e_2	0	3	3	+
e_3	3	0	0	+
e_4	2	3	0	+
e_5	4	3	0	-
e_6	4	0	3	-
e_7	3	0	0	-
e_8	0	0	0	-

\ominus

$e_5 : a \ b \ a \ b \ a \ b \ a \ d$

$e_6 : a \ c \ a \ c \ a \ c \ a \ d$

$e_7 : e \ a \ e \ d \ a \ e \ a$

$e_8 : d \ e \ d \ e \ d \ e \ d$

MULTIVARIATE BERNoulli
Bit Vector

	a	b	c	class
e_1	0	1	0	+
e_2	0	1	1	+
e_3	1	0	0	+
e_4	1	1	0	+
e_5	1	1	0	-
e_6	1	0	1	-
e_7	1	0	0	-
e_8	0	0	0	-

Smoothing

MULTINOMIAL
Count Vector

	a	b	c	class
e ₁	0	3	0	+
e ₂	0	3	3	+
e ₃	3	0	0	+
e ₄	2	3	0	+
v ₁	1	0	0	+
v ₂	0	1	0	+
v ₃	0	0	1	+
e ₅	4	3	0	-
e ₆	4	0	3	-
e ₇	3	0	0	-
e ₈	0	0	0	-
v ₁	1	0	0	-
v ₂	0	1	0	-
v ₃	0	0	1	-

MULTIVARIATE BERNoulli
Bit Vector

	a	b	c	class
e ₁	0	1	0	+
e ₂	0	1	1	+
e ₃	1	0	0	+
e ₄	1	1	0	+
v ₁	1	1	1	+
v ₂	0	0	0	+
e ₅	1	1	0	-
e ₆	1	0	1	-
e ₇	1	0	0	-
e ₈	0	0	0	-
v ₁	1	1	1	-
v ₂	0	0	0	-

Answer

Answer

Bernoulli

"event with two possible outcomes"

θ prob "success"

$$\text{mean } \mathbb{E}[X] = \theta \quad \text{variance } \mathbb{E}[(X - \mathbb{E}[X])^2] = \theta(1-\theta)$$

Binomial

"number of successes s in n trials" (prob θ)

$$P(s, n) = \binom{n}{s} \theta^s (1-\theta)^{n-s}$$

$$\text{mean } \mathbb{E}[s] = n\theta \quad \text{variance } \mathbb{E}[(s - \mathbb{E}[s])^2] = n\theta(1-\theta)$$

Categorical

"event with > 2 possible outcomes"

aka
Generalized
Bernoulli

or
Discrete

"number of k -categorical outcomes in n iid trials" (prob $\vec{\theta}$)

$\vec{X} = (X_1, X_2, \dots, X_k)$ is a k -vector of counts

$$P(\vec{X} = (x_1, x_2, \dots, x_k)) =$$

$$n! \frac{\theta_1^{x_1}}{x_1!} \cdots \frac{\theta_k^{x_k}}{x_k!} \quad \text{with } \sum_{i=1}^k x_i = n$$

Dirichlet
is
conjugate
prior

Answer

Raw data - text, w/ words (emails, Tweets, etc)				
				(+)
e ₁ :	b	d	e	bb de
e ₂ :	b	c	e	bb dd e cc
e ₃ :	a	d	a	d e a ee
e ₄ :	b	a	d	bed ab
				(-)
e ₅ :	a	b	a	b ab a ed
e ₆ :	a	c	a	c a c a c a ed
e ₇ :	e	a	e	a e d a e a
e ₈ :	d	e	d	e d e d
				(-)
<u>Counts</u>				
	a	b	c	class
e ₁ :	0	3	0	+
e ₂ :	0	3	3	+
e ₃ :	3	0	0	+
e ₄ :	2	3	0	+
e ₅ :	4	3	0	-
e ₆ :	4	0	3	-
e ₇ :	3	0	0	-
e ₈ :	0	0	0	-
<u>Big vec</u>				
	a	b	c	class
e ₁ :	0	1	0	+
e ₂ :	0	1	1	+
e ₃ :	1	0	0	+
e ₄ :	1	1	0	+
e ₅ :	1	1	0	-
e ₆ :	1	0	1	-
e ₇ :	1	0	0	-
e ₈ :	0	0	0	-
<u>Sums</u>	(5 9 3)	(+)	<u>Sums</u>	(2 3 1) (+)
	(11 3 3)	(-)		(3 1 1) (-)
<u>Smooth</u> (add 1 word)	(6 10 4)	(+)	<u>Smooth</u> Laplace *	[2 pseudo-documents: (111..) & (000..)]
	($\frac{12}{20}$ $\frac{4}{20}$ $\frac{4}{20}$)	(-)		($\frac{3}{6}$ $\frac{4}{6}$ $\frac{3}{6}$) (+)
	(0.3 0.5 0.2)	(+)		($\frac{4}{6}$ $\frac{2}{6}$ $\frac{2}{6}$) (-)
	(0.6 0.2 0.2)	(-)		(0.5 0.67 0.33) (+)
				(0.67 0.33 0.33) (-)
DATA:	multivariate Bernoulli	(x_1, x_2, x_3, \dots)	multinomial (count vector)	(x_1, x_2, x_3, \dots)

Answer

Prediction (using multi-variate Bernoulli)

	a	b	c	class
Model params.	(0.5	0.67	0.33)	⊕
	(0.67	0.33	0.33)	⊖

New instance: $\vec{x} = (1 \quad 1 \quad 0)$

Hinc $\underset{\text{class } \in \{\oplus, \ominus\}}{\operatorname{arg\,max}} P(\vec{x} | \text{class})$

$$P(\vec{x} | \oplus) = 0.5 * 0.67 * (1 - 0.33) = 0.222$$

$$P(\vec{x} | \ominus) = 0.67 * 0.33 * (1 - 0.33) = 0.148$$

\hookrightarrow predict \oplus

$$\text{LR} = \frac{P(\vec{x} | \oplus)}{P(\vec{x} | \ominus)} = \frac{0.5}{0.67} * \frac{0.67}{0.33} * \frac{1 - 0.33}{1 - 0.33} = 3/2 > 1$$

Hinc predict \oplus if prior odds $\frac{P(\oplus)}{P(\ominus)} > 2/3$!