

COMP9417 Machine Learning and Data Mining

Lectures: Classification (2)

Topic: Questions from lecture

Version: with answers

Last revision: Fri 19 Jun 2020 16:22:49 AEST

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 22–26 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.

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

$$\text{First test} = \oplus \quad [\arg\max_h P(h|D) = P(D|h)P(h)]$$

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

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

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

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

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

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

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

Question 2 Work through the document classification examples on slides 57–75. 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.

Answer

Question 6 :

(+)

$e_1 : b d e b b d e$
 $e_2 : b c e b b d d e c c$
 $e_3 : a d a d e a e e$
 $e_4 : b a d b e d a b$

$e_5 : a b a b a b a e d$
 $e_6 : a c a c a c a e d$
 $e_7 : e a e d a e a$
 $e_8 : d e d e d$

(-)

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	-

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	-

Answer

Smoothing

MULTINOMIAL
Count Vector

	a	b	c	class
e_1	0	0	0	+
e_2	0	3	0	+
e_3	3	0	0	+
e_4	2	3	0	+
v_1	0	-1	0	+
v_2	0	0	0	+
v_3	-1	0	0	+
e_5	4	3	0	+
e_6	4	0	0	+
e_7	3	0	0	+
e_8	0	0	0	+
v_1	0	0	0	+
v_2	0	0	0	+
v_3	0	0	0	+

MULTIVARIATE BERNoulli
Bit Vector

	a	b	c	class
e_1	0	1	0	+
e_2	0	0	1	+
e_3	1	0	0	+
e_4	1	1	0	+
v_1	0	-1	0	+
v_2	0	0	0	+
e_5	1	1	1	+
e_6	0	0	0	+
e_7	1	-1	0	+
e_8	0	0	0	+
v_1	0	0	0	+
v_2	0	0	0	+
v_3	0	0	0	+

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" ($\text{prob } \theta$)

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

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

Categorical

"event with > 2 possible outcomes"

aka
Generalized
Bernoulli

or
Discrete

$\vec{\theta}$ prob of k possible outcomes, $\sum_{i=1}^k \theta_i = 1$

Multinomial

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

$\vec{X} = (x_1, x_2, \dots, x_k)$ is a k -vector of counts

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

$$\frac{n!}{x_1! x_2! \dots x_k!} \frac{\theta_1^{x_1}}{x_1!} \dots \frac{\theta_k^{x_k}}{x_k!} \quad \text{with } \sum_{i=1}^k x_i = n$$

Dirichlet
is
conjugate
prior

Answer

Raw data - text, w/ words (emails, Tweets, etc)

⊕
 $e_1: b \ a \ e \ bb \ de$
 $e_2: b \ c \ e \ bb \ dd \ e \ cc$
 $e_3: a \ d \ a \ d \ e \ a \ ee$
 $e_4: b \ a \ d \ b \ ed \ ab$

$$\theta_i = \frac{d + p_i m}{n + m}$$

$$\theta_i = \frac{d + \text{pseudo-counts}}{n + \text{pseudo-counts}}$$

$$\theta_i = \frac{d(\# \text{occ. doc} + 1)}{n(\# \text{doc} + 2)}$$

	Counts			class
	a	b	c	
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	-

	Bit vec			class
	a	b	c	
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	-

Sum $(5 \ 9 \ 3) \oplus$
 $(11 \ 3 \ 3) \ominus$

Sum $(2 \ 3 \ 1) \oplus$
 $(3 \ 1 \ 1) \ominus$

Smooth
 (add 1 word) $(6 \ 10 \ 4) \oplus$
 $(\frac{12}{20} \ \frac{4}{20} \ \frac{4}{20}) \ominus$
 $(0.3 \ 0.5 \ 0.2) \oplus$
 $(0.6 \ 0.2 \ 0.2) \ominus$

Smooth Laplace * [2 pseudo-documents: (1111...) & (0000...)]
 $(\frac{3}{6} \ \frac{4}{6} \ \frac{2}{6}) \oplus$
 $(\frac{4}{6} \ \frac{2}{6} \ \frac{2}{6}) \ominus$
 $(0.5 \ 0.67 \ 0.33) \oplus$
 $(0.67 \ 0.33 \ 0.33) \ominus$

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
Model params.	(0.5	0.67	0.33)
	⊕		class
	(0.67	0.33	0.33)
			⊖

New instance: $\vec{x} = \begin{pmatrix} a \\ 1 \\ 1 \\ 0 \end{pmatrix}$

$$\underline{H_{ML}} \quad \underset{\text{class } c \in \{\oplus, \ominus\}}{\operatorname{argmax}} \quad P(\vec{x} | \text{class } c)$$

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

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

→ predict ⊕

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

$$\underline{H_{MAP}} \quad \text{predict } \oplus \text{ if prior odds } \frac{P(\oplus)}{P(\ominus)} > 2/3$$

