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 $\neg$*cancer* following these two tests? Note: you can assume that the two tests are independent.

Answer

Assume: two tests are independent.

Priors $P(\text{cancer}) = 0.008$ $\quad P(\neg\text{cancer}) = 0.992$

Likelihoods $P(\oplus|\text{cancer}) = 0.98$ $\quad P(\ominus|\text{cancer}) = 0.02$ $\quad$ Given

$P(\oplus|\neg\text{cancer}) = 0.03$ $\quad P(\ominus|\neg\text{cancer}) = 0.97$

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

$$P(\oplus|\text{cancer})\,P(\text{cancer}) = 0.98 \times 0.008 = 0.0078$$
$$P(\oplus|\neg\text{cancer})\,P(\neg\text{cancer}) = 0.03 \times 0.992 = 0.0298$$

$\Rightarrow$ MAP = $\neg$cancer

Second test = $\oplus$ $\quad D = \{\overset{d_1}{\oplus}, \overset{d_2}{\oplus}\}$

$P(d_1|h)\,P(d_2|h)\,P(h)$ // By independence of data!

$$P(\oplus|c)\,P(\oplus|c)\,P(c) = 0.98 \times 0.98 \times 0.008 = 0.0077$$
$$P(\oplus|\neg c)\,P(\oplus|\neg c)\,P(\neg c) = 0.03 \times 0.03 \times 0.992 = 0.0009$$

$\Rightarrow$ MAP = 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.

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	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

Bernoulli "event with two possible outcomes"

θ prob^{ty} success

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

Binomial "number of successes s in n iid trials" (prob^{ty} θ)

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

mean $\mathbb{E}[s] = n\theta$ 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^{ty} of k possible outcomes, $\sum_{i=1}^{k} \theta_i = 1$

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

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

$$p(\vec{X} = (x_1, x_2, \ldots, x_k)) = n! \, \frac{\theta_1^{x_1}}{x_1!} \cdots \frac{\theta_k^{x_k}}{x_k!} \quad \text{with } \sum_{i=1}^{k} x_i = n$$

$\boxed{\text{Dirichlet is conjugate prior}}$

Answer

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

P_i cat. i

m-est. $\dfrac{d + p_i m}{n + m}$

⊕	⊖
e_1: b d e bb de	e_5: a b ab ab aed
e_2: b c e bb dd e cc	e_6: a c ac ac aed
e_3: a d a d e a ee	e_7: e a e d a e a
e_4: b a d bed ab	e_8: d e d ed

"pseudo-counts"

* Laplace

$\dfrac{d}{n}\left(\dfrac{\#\text{ occ. doc} + 1}{\#\text{ doc} + 2}\right)$

Counts

	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	–

Sum

(5 9 3) ⊕

(11 3 3) ⊖

Smooth (add 1 word)

(6 10 4) ⊕

$\left(\frac{12}{20}\ \frac{4}{20}\ \frac{4}{20}\right)$ ⊖

(0.3 0.5 0.2) ⊕

(0.6 0.2 0.2) ⊖

Bit vec

	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	–

Sum

(2 3 1) ⊕

(3 1 1) ⊖

Smooth * Laplace

[2 pseudo-documents: (1 1 1 ..) & (0 0 0 ..)]

$\left(\frac{3}{6}\ \frac{4}{6}\ \frac{2}{6}\right)$ ⊕

$\left(\frac{4}{6}\ \frac{2}{6}\ \frac{2}{6}\right)$ ⊖

(0.5 0.67 0.33) ⊕

(0.67 0.33 0.33) ⊖

DATA: multivariate Bernoulli $(X_1, X_2, X_3 \ldots)$ with $\theta_1\ \theta_2\ \theta_3 \ldots$; multinomial (count vector) $(X_1\ X_2\ X_3 \ldots)$

Answer

Prediction (using multi-variate Bernoulli)

Model parms.

$$\begin{array}{ccccc} a & b & c & \text{class} \\ (0.5 & 0.67 & 0.33) & \oplus \\ (0.67 & 0.33 & 0.33) & \ominus \end{array}$$

New instance $\vec{x}$:

$$\begin{array}{ccc} a & b & c \\ (1 & 1 & 0) \end{array}$$

H_{ML} $\underset{class \in \{\oplus, \ominus\}}{\arg\max} \; p(\vec{x} \mid class)$

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

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

$\hookrightarrow$ predict $\oplus$

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

H_{MAP} predict $\oplus$ if prior odds $\frac{p(\oplus)}{p(\ominus)} > 2/3$!