School of Computing and Information Systems
The University of Melbourne
COMP30027 MACHINE LEARNING (Semester 1, 2019)

Tutorial exercises: Week 3

Given the following dataset:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
TRAINING INSTANCES					
A	s	h	n	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N
TEST INSTANCES					
G	o	m	n	T	?
H	?	h	?	F	?

1. Build a probabilistic **model** based around the given training instances:
   (a) Calculate the **prior** probability $P(Outl = \texttt{s})$. Calculate the prior probabilities of the other attribute values in this data.
   (b) Find the **entropy** of (the distribution of the attribute values) for each of the six attributes, given this probabilistic model.
   (c) Calculate the **joint** probability $P(Outl = \texttt{s} \cap Temp = \texttt{h})$. Calculate some other joint probabilities, for pairs of attribute values from different attributes.
   (d) Calculate the **conditional** probability $P(Outl = \texttt{s} | Temp = \texttt{h})$. Calculate some other conditional probabilities.
2. Ensure that you can derive the **Naive Bayes** formulation.
3. Using the probabilistic model that you developed above, classify the test instances according to the method of **Naive Bayes**.
   (a) Using the "epsilon" smoothing method.
   (b) Using "Laplace" smoothing.

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
TRAINING INSTANCES					
A	s	h	n	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N
TEST INSTANCES					
G	o	m	n	T	?
H	?	h	?	F	?

1. Build a probabilistic **model** based around the given training instances:
   (a) Calculate the **prior** probability $P(Outl = \mathtt{s})$. Calculate the prior probabilities of the other attribute values in this data.
   (b) Find the **entropy** of (the distribution of the attribute values) for each of the six attributes, given this probabilistic model.
   (c) Calculate the **joint** probability $P(Outl = \mathtt{s} \cap Temp = \mathtt{h})$. Calculate some other joint probabilities, for pairs of attribute values from different attributes.
   (d) Calculate the **conditional** probability $P(Outl = \mathtt{s} | Temp = \mathtt{h})$. Calculate some other conditional probabilities.

1. (a) $P(Outl = s) = \frac{2}{6}$

(b) Entropy (bits) =

x是每个attribute name

$$H(X) = -\sum_{x \in X} P(x) \log_2 P(x)$$

针对每个 attribute column

$$H(Outl) = -\left[P(Outl=s)\log_2 P(Outl=s) + \cdots\right]$$

≈ 1.46 bits 注意单位

What is entropy?

信息所含的不确定性大小的度量

信息包含信息量少 → $H(x) = 0$

e.g. 太阳升起：{东方，西方}

P(东方) = 0.99999, P(西方) = 0.00001

$H(x) \approx 0$.

即包含事件的不确定性越大 → 信息越多

(c) Joint Probability : $P(A \cap B)$ 多件事同时发生的概率

$$P(Outl = s \cap Temp = h) = \frac{2}{6}$$

(d) Conditional Probability : $P(A|B) = \frac{P(A \cap B)}{P(B)}$

$P(Outl = s | Temp = h) =$

2. Ensure that you can derive the **Naive Bayes** formulation.

Naive Bayes formula:

$$\hat{C} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_i P(x_i | c_j)$$

---

3. Using the probabilistic model that you developed above, classify the test instances according to the method of **Naive Bayes**.

   (a) Using the "epsilon" smoothing method.

   (b) Using "Laplace" smoothing.

ID	Outl	Temp	Humi	Wind	PLAY
TRAINING INSTANCES					
A	s	h	n	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N
TEST INSTANCES					
G	o	m	n	T	?
H	?	h	?	F	?

"epsilon" smoothing:

$$P(x|c_j) = \frac{1}{a} \times 0 \times 0 \times \frac{1}{b} \times \frac{1}{c} = 0$$

$\Rightarrow$ introduce $\varepsilon \ll \frac{1}{\text{instance number}}$

$$P(x|c_j) = \frac{1}{a} \times \varepsilon \times \varepsilon \times \frac{1}{b} \times \frac{1}{c} = \frac{1}{abc} \times \varepsilon^2$$

(a) test G:

分别算 c_1, c_2 (Y|N)

$$Y = P(G_x|Y) = P(Y)P(o|Y)P(m|Y)P(n|Y)P(T|Y)$$

$$= \frac{3}{6} \times \boxed{\frac{0}{3}} \times \boxed{\frac{0}{3}} \times \frac{2}{3} \times \frac{2}{3} = 0$$

$$= \frac{3}{6} \times \varepsilon \times \varepsilon \times \frac{2}{3} \times \frac{2}{3} = \frac{2\varepsilon^2}{9}$$

$$N = P(G_x|N) = P(N)P(o|N)P(m|N)P(n|N)P(T|N)$$

$$= \frac{3}{6} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \varepsilon = \frac{\varepsilon}{54}$$ ✓

当出现 missing value (?), NB 直接忽略,

test H:

$$Y: P(H_x|Y) = \frac{3}{6} \times \frac{3}{3} \times \frac{1}{3} = \frac{1}{9}$$

$$N: P(H_x|N) = \frac{3}{6} \times \frac{1}{3} \times \frac{3}{3} = \frac{1}{6}$$ ✓

Using "Laplace" smoothing.

$$\hat{P}_L(a|c) = \frac{1 + freq(a,c)}{|V| + freq(C)}$$

$|V|$: # of attribute value

(b) test G:

Y: $P(G_x|Y) = P(Y)\ \hat{P}_L(o|Y)\ \hat{P}_L(m|Y)\ \hat{P}_L(n|Y)\ \hat{P}_L(T|Y)$

$$= \frac{3}{6} \times \frac{0+1}{3+3} \times \frac{0+1}{3+3} \times \frac{1+2}{2+3} \times \frac{1+2}{2+3} = 0.005$$ ✓

N: $P(G_x|N) = 0.0044$

test H:

Y: $P(H_x|Y) = \frac{3}{6} \times \frac{1+2}{3+3} \times \frac{1+1}{2+3} = 0.01$

N: $P(H_x|N) = \frac{3}{6} \times \frac{1+1}{3+3} \times \frac{1+3}{3+3} = 0.013$ ✓

Different smoothing method Different Result.