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.

(a) $P(\text{outl} = s) = \frac{2}{6} = \frac{1}{3}$

(b) entropy → uncertainty

$H(x) = 0$ → less information → certainty.

$H(x) = 1$ → more information → uncertainty.

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

attribute name

attribute value.

e.g. $H(\text{Outl}) = -\left(P(\text{outl}=s)\log_2 P(\text{outl}=s) + P(\text{outl}=o)\log_2 P(\text{outl}=o) + \cdots\right)$

$= -\left(\frac{2}{6}\log_2\frac{2}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{3}{6}\log_2\frac{3}{6}\right)$

$= 1.46$ bits.

(c) Joint probability $P(A \cap B)$ ⇒ 同时发生

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

(d) Conditional probability $P(A|B)$ ⇒ given B 发生 A

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(\text{Outl} = s | \text{Temp} = h) = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$

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

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

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	?

(a) "Epsilon": $0 \rightarrow \varepsilon \ll \frac{1}{N}$ ↖ instance Num

test G:

Class = N: $\frac{3}{6} \times (\varepsilon \times \varepsilon \times \frac{2}{3} \times \frac{2}{3}) = \frac{2}{9}\varepsilon^2$

Class = Y: $\frac{3}{6} \times (\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \varepsilon) = \frac{1}{54}\varepsilon$ ✓

test H: (当出现 missing value 时, ignore it)

class = N: $\frac{3}{6} \times (\frac{1}{3} \times \frac{3}{3}) = \frac{1}{6}$ ✓

class = Y: $\frac{3}{6} \times (\frac{2}{3} \times \frac{1}{3}) = \frac{1}{9}$

(b) "Laplace": $\hat{P}_{\mathcal{L}}(a|c) = \frac{1 + freq(a,c)}{|V| + freq(c)}$

↑ # of attribute value type.

test G:

Class = N: $\frac{3}{6} \times (\frac{1+0}{3+3} \times \frac{1+0}{3+3} \times \frac{1+2}{2+3} \times \frac{2+2}{2+3}) = 0.005$ ✓

Class = Y: $\frac{3}{6} \times (\frac{1+1}{3+3} \times \frac{1+1}{3+3} \times \frac{1+1}{2+3} \times \frac{1+0}{2+3}) = 0.0044$

test H:

Class = N $\frac{3}{6} \times (\frac{1+2}{3+3} \times \frac{1+1}{2+3}) = 0.01$

Class = Y $\frac{3}{6} \times (\frac{1+1}{3+3} \times \frac{1+3}{3+3}) = 0.013$ ✓

Different smothing method, different result