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

Tutorial exercises: Week 6

ID	A (°C)	B (mm)	C (hPa)	CLASS
1	22.5	4.6	1021.2	AUT
2	16.7	21.6	1027.0	AUT
3	29.6	0.0	1012.5	SUM
4	33.0	0.0	1010.4	SUM
5	13.2	16.4	1019.5	SPR
6	14.9	8.6	1016.4	SPR
7	18.3	7.8	995.4	WIN
8	16.0	5.6	1012.8	WIN

1. What is **Discretisation**, and where might it be used?
   (a) Summarise some approaches to **supervised** discretisation.
   (b) Discretise the above dataset according to the (unsupervised) methods of **equal width**, **equal frequency**, and **k-means** (breaking ties where necessary).
2. Find the (sample) **mean** and (sample) **standard deviation**[1] for the attibutes in the above dataset:
   (a) In its entirety, and;
   (b) For each individual class[2].
   (c) How could we use this information when building a classifier over this data?

   Given the following dataset:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
A	s	h	h	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

3. If we wished to perform **feature selection** (or **feature weighting**) on this dataset, where the class is PLAY:
   (a) Which of *Humi* and *Wind* has the greatest **Pointwise Mutual Information** for the class Y? What about N?
   (b) Which of the attributes has the greatest **Mutual Information** for the class, as a whole? (Note that we need to extend the lecture definition to handle non–binary attributes.)

[1] n.b. You might need a calculator.
[2] We would ideally do this with more instances!

ID	A (°C)	B (mm)	C (hPa)	CLASS
1	22.5	4.6	1021.2	AUT
2	16.7	21.6	1027.0	AUT
3	29.6	0.0	1012.5	SUM
4	33.0	0.0	1010.4	SUM
5	13.2	16.4	1019.5	SPR
6	14.9	8.6	1016.4	SPR
7	18.3	7.8	995.4	WIN
8	16.0	5.6	1012.8	WIN

1. What is **Discretisation**, and where might it be used?
   (a) Summarise some approaches to **supervised** discretisation.
   (b) Discretise the above dataset according to the (unsupervised) methods of **equal width**, **equal frequency**, and **k-means** (breaking ties where necessary).

Discretisation : continuous → nominal. (attribute)

When we have a discrete classifier, but the data is continuous.

(a) Sort the values, and create nominal value for a region where most of the instances having the same label.

(b) i) Equal width. (Based on value range, regardless number)

$$\text{Bin Range} = \frac{\max - \min}{n}$$

ii) Equal frequency (Based on number of instance in each bin)

$$\text{Bin Range} = \frac{\text{total number}}{n}.$$

Two above should SORT the attribute value in former.

iii) K-means.

K seeds → stable K clusters → give each cluster a bin name,

2. Find the (sample) **mean** and (sample) **standard deviation**[1] for the attibutes in the above dataset:

   (a) In its entirety, and;

   (b) For each individual class[2].

   (c) How could we use this information when building a classifier over this data?

Given the following dataset:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
A	s	h	h	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

2.(a) mean : $\mu_A = \frac{1}{N} \sum A_i$

Standard Deviation : $\sigma_c = \sqrt{\frac{\sum (C_i - \mu_c)^2}{(N-1)}}$

For Attribute C : $\mu_c = \frac{1}{N} \sum C_i = \frac{1}{8} (1012.2 + 1027.0 + \dots)$

$= 1014.4$

$\sigma_c = \sqrt{\frac{(1012.2 - 1014.4)^2 + \dots}{7}}$

$= 9.40$

(b) Same as a)

(c) We can build a normal pdf. which allow us estimate the probability of observing any given value. using pointwise estimation.

3. If we wished to perform **feature selection** (or **feature weighting**) on this dataset, where the class is PLAY:

   (a) Which of *Humi* and *Wind* has the greatest **Pointwise Mutual Information** for the class Y? What about N?

   (b) Which of the attributes has the greatest **Mutual Information** for the class, as a whole? (Note that we need to extend the lecture definition to handle non–binary attributes.)

3. (a) $PMI(A;C) = \log_2 \frac{P(A \cap C)}{P(A)P(C)}$ ← only for binary attributes and classes.

$PMI(Humi;Y) = \log_2 \frac{\frac{2}{6}}{\frac{4}{6} \times \frac{3}{6}} = \log_2(1) = 0$ uncorrelated

$PMI(Wind;Y) = \log_2 \frac{\frac{0}{6}}{\frac{2}{6} \times \frac{3}{6}} = \log_2(0) = -\infty$ perfectly negative correlated

(b) $MI(X;C) = \sum_{x \in X} \sum_{c \in \{Y,N\}} P(x,c)\, PMI(x;c)$

For Outl :

$$MI = P(s,Y)PMI(s;Y) + P(s,N)PMI(s;N) +$$
$$P(o;Y)PMI(o;Y) + P(o,N)PMI(o;N) +$$
$$P(r;Y)PMI(r;Y) + P(r,N)PMI(r;N).$$

$$= 0\log_2 0 + \cdots$$

$$= 0.541$$