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

Continuous : {1, 2, 3, 4, 5, 6, ...}

nominal : {Sunny, rainy, overcast}

Ordinal : {low, med, height}

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).

1. Discretisation : we have continuous (numeric) attribute, but we wish to have a nomial (or ordinal) attribute

when use? when leaner is a nominal

(a) Sort the possible value, create range, map each range a discrete name

(b) equal width :

1. find largest, smallest, chose n-value

2. $\frac{\text{difference}}{n}$ be range

equal frequency :

1. Sort, Choose n-value

2. $\frac{\text{\# of instance}}{n}$ be number of range

K-means : (clustering)

1. Choose K points as K seeds

2. iterate

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) $\mu_c = \frac{1}{N}\sum C_i = \frac{1}{8}(1021.2 + 1027.0 + \cdots)$

$= 1014.4$

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

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.)

Point wise Mutual Information.

$PMI(A;C) = \log_2 \frac{P(A\cap C)}{P(A)P(C)}$

attribute with greatest PMI.

↓

best feature (correlated to class)

$PMI(Humi;Y) = \log_2 \frac{P(Humi \cap Y)}{P(Humi)P(Y)}$

$= \log_2 \frac{\frac{2}{6}}{\frac{4}{6} \times \frac{3}{6}} = 0$ ← non-correlated

$PMI(Wind;Y) = \log_2 \frac{0}{-} = \log_2 0 = -\infty$

negative correlated

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

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

$= 0.541$ - best attribute

$MI(temp) = 0.110$. — not very good

$MI(Humi) = 0$ - unhelpful

$MI(wind) = 0.459$