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

Tutorial exercises: Week 4

1. Consider the following 10 instances, given so-called "gold standard" labels (assuming a 3-class problem), and the output of four supervised machine learning models:

Instance	Gold	1	2	3	4
1	A	A	A or B	A	A
2	B	A	B or C	A	?
3	A	A	A	A	A
4	C	C	B or C	A	?
5	B	B	A or B or C	A	?
6	C	A	A or C	A	?
7	C	A	A or B or C	A	?
8	A	C	A or B	A	A
9	A	A	A	A	?
10	A	A	A or C	A	A

   (a) Where possible, calculate the **accuracy** and **error rate** of the four models.

   (b) Where possible, calculate the **precision** and **recall**, treating class A as the "positive" class. Do the same for the B and C classes, in turn, and then calculate the **macro-averaged precision and recall**.

2. What is the difference between evaluating using a **holdout** strategy and evaluating using a **cross–validation strategy**?

   (a) What are some reasons we would prefer one strategy over the other?

3. For the following dataset:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
TRAINING INSTANCES					
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
TEST INSTANCES					
G	o	c	n	T	?
H	s	m	h	F	?

   (a) Classify the test instances using the method of 0-R.

   (b) Classify the test instances using the method of 1-R.

   (c) Classify the test instances using the ID3 **Decision Tree** method:

      i. Using the **Information Gain** as a splitting criterion

      ii. Using the **Gain Ratio** as a splitting criterion

1. Consider the following 10 instances, given so-called "gold standard" labels (assuming a 3-class problem), and the output of four supervised machine learning models:

Instance	Gold	1	2	3	4
1	A	A	A or B	A	A
2	B	A	B or C	A	?
3	A	A	A	A	A
4	C	C	B or C	A	?
5	B	B	A or B or C	A	?
6	C	A	A or C	A	?
7	C	A	A or B or C	A	?
8	A	C	A or B	A	A
9	A	A	A	A	?
10	A	A	A or C	A	A

multi-class

(a) Where possible, calculate the **accuracy** and **error rate** of the four models.

(b) Where possible, calculate the **precision** and **recall**, treating class A as the "positive" class. Do the same for the B and C classes, in turn, and then calculate the **macro-averaged precision and recall**.

(a) Formal Formula : $Acc = \frac{TP + TN}{TP + FP + TN + FN}$

$ER = 1 - Acc$

i) $acc = \frac{6}{10}$ $ER = 1 - \frac{6}{10} = \frac{4}{10}$

ii) $acc = \frac{TP = 10 + TN = 0}{10 + FP = 10 + 0 + FN = 0} = \frac{10}{20}$ $ER = 1 - \frac{10}{20} = \frac{10}{20}$

iii) $acc = \frac{5}{10}$ $ER = \frac{5}{10}$

iv) $acc = \frac{TP = 4 + TN = 0}{4 + FP = 0 + 0 + FN = 6} = \frac{4}{10}$ $ER = 1 - \frac{4}{10} = \frac{6}{10}$

(b) Pricison $= \frac{TP}{TP + FP}$ Recall $= \frac{TP}{TP + FN}$

Assume A as interesting class .

i) $P = \frac{4}{7}$ $R = \frac{4}{5}$

ii) $P = \frac{5}{8}$ $R = \frac{5}{5}$

iii) $P = \frac{5}{10}$ $R = \frac{4}{5}$

iv) $P = \frac{4}{4}$ $R = \frac{4}{5}$

macro - averaging P & R :

$P_m = \frac{1}{n} \sum_i P_i$

$R_m = \frac{1}{n} \sum_i R_i$

2. What is the difference between evaluating using a **holdout** strategy and evaluating using a **cross–validation strategy**?

   (a) What are some reasons we would prefer one strategy over the other?

hold-out : partition data into train and test set, build the model on former,

evaluate the model later.

Cross-validation : do the same above, but we partition the data set into several parts. and iterate multiple times. Let each partition as test set and rest as train set.

as prefer cross-validation

For hold out :

1. random variation, depending on which data split to train, which to test set.

2. Any instance in train set is excluded in test set. This means our estimate of performance may way off.

3. Result may change a lot

For cross-validation;

1. Cross-validation solve this question by averaging the values.

2. each instance is used for testing, but appears in train set in other partitions.

3. Take a long time.

3. For the following dataset:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
TRAINING INSTANCES					
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
TEST INSTANCES					
G	o	c	n	T	?
H	s	m	h	F	?

(a) Classify the test instances using the method of 0-R.

(b) Classify the test instances using the method of 1-R.

(c) Classify the test instances using the ID3 **Decision Tree** method:

i. Using the **Information Gain** as a splitting criterion

ii. Using the **Gain Ratio** as a splitting criterion

0 - R : majority class.

1 - R : choose attribute with the smallest error rate in tree stump. (simiply counting error made in trainning set)

(a) N (need break - tie which is better)

				error
(b) For Outl	= s :	2 s	2 N	0
	= o :	1 o	1 N	0
	= r :	3 r	2 Y 1 N	1

total = 1

- - - - find smallest error attribute

(c) Classify the test instances using the ID3 **Decision Tree** method:

i. Using the **Information Gain** as a splitting criterion

ii. Using the **Gain Ratio** as a splitting criterion

(c) Information Gain: the difference between the entropy of the parent node, and the average entropy across its daughter node. (mean information)

$$IG(A|R) = H(R) - \sum_{i \in A} P(A=i) \underbrace{H(A=i)}$$

看 label 的分布

For Root: $H(R) = -(\frac{3}{6}\log_2\frac{3}{6} \times 2) = 1$　　一般 2 labels

For Outl: $H(Outl=s) = -(1\log_2 1 + 0\log_2 0) = 0$

$H(Outl=o) = -(1\log_2 1 + 0\log_2 0) = 0$

$H(Outl=r) = -(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}) = 0.9183.$

$$\text{Mutual Information} = \sum_{i=1}^{M} P(x_i) H(x_i)$$

$MI(s,o,r) = \frac{2}{6}(0) + \frac{1}{6}(0) + \frac{3}{6}(0.9183) = 0.4592$

$IG(Outl|R) = H(R) - MI(s,o,r) = 1 - 0.4592 = 0.5408.$

注：算所有 columns 过后. ID has the best IG, but each daughter is purely of a single class - However we could get an useless classifier, since ID is unique.

ii) Gain Ratio.

$$\frac{\# i}{\text{total instance}}$$

$$GR(A) = \frac{IG(A)}{SI(A)}$$

split information. $SI(A) = -\sum_{i \in A} P(A=i) \log_2 P(A=i)$

For Outl : $SI(Outl) = -\left(\frac{2}{6}\log_2\frac{2}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{3}{6}\log_2\frac{3}{6}\right) \approx 1.459$

$$GR(Outl) = \frac{0.5408}{1.459} \approx 0.3707$$

Procedure same as i) IG