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

Tutorial exercises: Week 9

1. Let's revisit the logic behind the **voting** method of classifier combination (used in **Bagging**, **Random Forests**, simple **Stacking**, and **Boosting** to some extent): Let's make a few assumptions (some of which we'll try to relax later):

   (1) We have a two–class problem;

   (2) The test (and training) instances are roughly evenly divided between the two classes;

   (3) Our classifiers predict the test instances roughly in proportion to the distribution of the classes;

   (4) We are building an ensemble out of two classifiers;

   (5) The errors between the two classifiers are **uncorrelated**.

   (a) First, let's assume our two classifiers both have an **error rate** of $e = 0.4$, calculated over 1000 instances.

      i. Build the **confusion matrix** for these classifiers, based on the assumptions above.

      ii. On the table overleaf, indicate the number of instances in the (count) column for the first two systems — a couple of values have been filled out; for example, there are 180 instances where the actual class is A, and both systems predicted A.

      iii. Assuming that the voting ties are broken randomly, what the the expected error rate of the voting ensemble?

   (b) What if we add a third classifier, also with error rate 0.4? Fill in the rest of the table, and determine the error rate of this ensemble. Why has adding a third system caused it to improve?

   (c) Now consider two classifiers, one (1) with $e_1 = 0.1$ and the second with $e_2 = 0.2$.

      i. Build the two confusion matrices.

      ii. Fill out the second table overleaf. Some values are given. Determine the expected error rate of the ensemble.

      iii. Add another system with $e_3 = 0.2$; does the error rate improve this time?

      iv. What if the errors between the systems were very highly correlated instead? What will happen to the error rate then? What do you think would happen if we added many more highly correlated classifiers to the ensemble?

   (d) (Extension) Find general forms for the rightmost values in the tables:

      i. for N instances and error rates $e_{1,2,3}$;

      ii. and, instead of the true labels being evenly divided between the two classes, a fraction α of the instances are class A, and $(1 - \alpha)$ are class B;

      iii. and, instead of the classifier making predictions in the ratio of the true labels, it is potentially biased, predicting class A for a fraction β of the instances, and $(1 - \beta)$ for class B [Hint: the A-A cell in the confusion matrix should be $\frac{N}{2}(\alpha + \beta - e)$]

   (e) Why can't we easily relax assumption (1) with the information given?

System 1 System 2 System 3

actual

	Predictions (all $e = 0.4$)				
	1	2	(count)	3	(count)
A	A	A	180	A	108
				B	72
		B	120	A	72
				B	48
	B	A	120	A	72
				B	48
		B	80	A	48
				B	32
B	A	A	80	A	32
				B	48
		B	120	A	48
				B	72
	B	A	120	A	48
				B	72
		B	180	A	72
				B	108

actual

System 1 sys 2 sys 3

	Predictions ($e_1 = 0.1$, $e_2, e_3 = 0.2$)				
	1	2	(count)	3	(count)
A	A	A	360	A	288
				B	72
		B	90	A	72
				B	18
	B	A	40	A	32
				B	8
		B	10	A	8
				B	2
B	A	A	10	A	2
				B	8
		B	40	A	8
				B	32
	B	A	90	A	18
				B	72
		B	360	A	72
				B	288

(a) First, let's assume our two classifiers both have an **error rate** of $e = 0.4$, calculated over 1000 instances.

i. Build the **confusion matrix** for these classifiers, based on the assumptions above.

ii. On the table overleaf, indicate the number of instances in the (count) column for the first two systems — a couple of values have been filled out; for example, there are 180 instances where the actual class is A, and both systems predicted A.

iii. Assuming that the voting ties are broken randomly, what the the expected error rate of the voting ensemble?

i.

$e = 0.4$		Actual T (A)	U (B)
predict	T	300	200
	U	200	300

error rate $= \frac{FP + FN}{TP + TF + FP + FN} = 0.4$

ii, page 2

iii, The esemble will make mistake when 2 classifiers label wrong at the same time, 80 A instances, both classified as B, 80 B instances both classified as A. total 160 errors.

240 instances A for one predict A, one predict B

240 instances B for one predict B. one predict A

the tie is broken, so total 480 instances and 240 instances wrong.

Total is 400 $\rightarrow \frac{400}{1000} = 0.4 \rightarrow$ error rate

(b) What if we add a third classifier, also with error rate 0.4? Fill in the rest of the table, and determine the error rate of this ensemble. Why has adding a third system caused it to improve?

We don't need to worry about breaking tie.

Only make mistake for 2of 3 or 3of3 make mistake.

$$\Big(\underbrace{48 \times 3}_{\text{2of3}} + \underbrace{32}_{\text{2of2}}\Big) \times 2 = (144 + 32) \times 2$$

$$= 352$$

$$e = \frac{352}{1000} = 35.2\% \text{ error rate}$$

found: even though 3rd system has the same 0.4 error rate.

but essemble impsoved.

Since the errors being uncorrelated: ← error position diff.

if error perfect correlated, we can see no improvement

if error mostly correlated, we can see a little improvement

(c) Now consider two classifiers, one (1) with $e_1 = 0.1$ and the second with $e_2 = 0.2$.

i. Build the two confusion matrices.

ii. Fill out the second table overleaf. Some values are given. Determine the expected error rate of the ensemble.

iii. Add another system with $e_3 = 0.2$; does the error rate improve this time?

iv. What if the errors between the systems were very highly correlated instead? What will happen to the error rate then? What do you think would happen if we added many more highly correlated classifiers to the ensemble?

i)

Actual

	$e = 0.1$	A	B	$e = 0.2$	A	B
Predict	A	450	50		400	100
	B	50	450		100	400

ii) $\left(\frac{90}{2} + \frac{40}{2} + 10\right) \times 2 = 150 \quad \frac{150}{1000} = 15\%$ error rate.

iii) $\left(\frac{36}{2} + \frac{16}{2} + \frac{16}{2} + 2\right) \times 2 = 72 \quad \frac{72}{1000} = 7.2\%$

the improvement happens even though c_3 is wore (0.2)

iv) if highly correlated, classifiers will make same mistake

no impovent on error rate.

if 2 is correlated, another one is uncorrelated.

2 > 1. Classifier 3 will be treated as error instance

if classifiers are highly corelated → no / less improvement