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

Tutorial sample solutions: 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.
      - Recall that a confusion matrix lists the number of instances where the classifier has predicted a certain class, and which were of some class.
      - The only confusion matrix consistent with our assumptions above, having 1000 instances and an error rate of 0.4 is as follows:

$e = 0.4$		Actual A	Actual B
Predicted	A	300	200
	B	200	300

      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.
      - See the final page.

      iii. Assuming that the voting ties are broken randomly, what the the expected error rate of the voting ensemble?
      - The ensemble will make mistakes when both classifiers produce the wrong label: there are 80 A instances, where both classifiers say B, and 80 B instances where both classifiers say A. This is 160 errors.
      - It will also make mistakes when the two classifiers disagree, and the tie is broken toward the wrong class; this is expected to happen for half of each set of instances where the classifiers disagree (because one of the two classes is wrong): there are 240 A instances where one of the classifiers says A and the other says B, and 240 B instances with the same disagreement. This is 480 instances in total, of which 240 will be errors.
      - Overall, there are expected to be 400 errors out of 1000 instances, so the expected error rate is (still) 40%.

   (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?
      - With three systems, we don't have to worry about breaking ties; the ensemble will make errors where 2 or 3 of the classifiers predict the wrong label.

- Reading off the table, there are $48 + 48 + 48 = 144$ A instances, where two classifiers predict B, as well as 32 A instances where all three classifiers predict B, which is 176 errors in total.
- Similarly, there are 176 erroneous B instances, which gives 352 errors in total, for an error rate of 35.2%.
- Even though the third system also has the same error rate, the ensemble has improved, because adding the third system has allowed us to disambiguate the instances where the first two classifiers were tied, better than random guessing.
- This relies on the errors being uncorrelated: if the errors were perfectly correlated, we would see no improvement; if the errors were mostly correlated, we would see only 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.
- Again, based on our assumptions above, we find unique solutions for the confusion matrices:

$e = 0.1$		Actual A	Actual B
Predicted	A	450	50
	B	50	450

$e = 0.2$		Actual A	Actual B
Predicted	A	400	100
	B	100	400

ii. Fill out the second table overleaf. Some values are given. Determine the expected error rate of the ensemble.
- See the final page for the table.
- This ensemble makes $45 + 20 + 10 + 10 + 20 + 45 = 150$ errors, which means that the error rate is 15% — worse than the first classifier by itself!

iii. Add another system with $e_3 = 0.2$; does the error rate improve this time?
- This ensemble makes $18 + 8 + 8 + 2 + 2 + 8 + 8 + 18 = 72$ errors, so the error rate is 7.2%; better than any classifier by itself. This improvement happens even though the classifier that we added has a worse accuracy that the ensemble!

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?
- Basically, all of the systems will be making the same predictions; so, the error rate will be roughly the same as the correlated classifiers, and voting is unlikely to improve the ensemble. Even if two of the classifiers are correlated, and the third is uncorrelated, the two correlated systems will tend to "out-vote" the third system on erroneous instances.
- Consequently, if we have many highly correlated classifiers, voting is likely to be useful only for a small number of instances.

(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?
- Basically, the error rate doesn't give us enough information about *what sorts of errors* the classifier is making in a multi–class problem. Consequently, we can't fill out the confusion matrices like we did in this problem.

- It is also difficult to make estimates where the ensemble is partly correct. For example, let's say we have a 4–class problem, and an ensemble of 5 classifiers. When 2 of the classifiers predict the right class and 3 of the classifiers predict the wrong class:
  - If the three wrong classifiers all choose the same (wrong) label, the ensemble will be wrong.
  - If two of the three wrong classifiers choose the same (wrong) label, but the third classifier chooses a different (wrong) label, there will be a tie; so the ensemble will be half–wrong.
  - If the three wrong classifiers all choose different (wrong) labels, the ensemble will actually be right!
- We can use combinatorics to estimate the likelihood of these situations, but the practical behaviour of ensembles in these situations tends to be very different to the theoretical behaviour.

		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

		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