School of Computing and Information Systems
The University of Melbourne
COMP30027 Machine Learning (Semester 1, 2021)

Tutorial: Week 8

1. How is **Logistic Regression** similar to **Naive Baye*s*** and how is it different? In what circumstances would the former be preferable, and in what circumstances would the latter?

2. Let's revisit the logic behind the voting method of classifier combination (used in Bagging, Random Forests, and Boosting to some extent). We are assuming that the errors between the two classifiers are uncorrelated

   (i). First, let's assume our three independent classifiers both have an error rate of e = 0.4, calculated over 1000 instances with binary labels (500 A and 500 B).

      a. Build the confusion matrices for these classifiers, based on the assumptions.

      b. Using that the majority voting, what the expected error rate of the voting

   (ii). Now consider three classifiers, first with $e_1 = 0.1$, the second and third with $e_2 = e_3 = 0.2$.

      a. Build the confusion matrices.

      b. Using the majority voting, what the expected error rate of the voting ensemble?

   (iii). What if we relax our assumption of independent errors? In other words, what will happen if the errors between the systems were very highly correlated instead? (Systems make similar mistakes.)

3. Given the following dataset, we wished to perform feature selection on this dataset, where the class is PLAY:

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

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

   (ii). Which of the attributes has the greatest *Mutual Information* for the class, as a whole?