

School of Computing and Information Systems
The University of Melbourne
COMP90049 Introduction to Machine Learning (Semester 1, 2023)

Week 11

1. What is **gradient descent**? Why is it important?
2. How is **Logistic Regression** similar to **Naive Bayes** and how is it different? In what circumstances would the former be preferable, and in what circumstances would the latter?
3. Describe how to build Random Forrest for a given data?
 - (a) What is the benefit of bagging?
 - (b) What is the impact of the size of the number of trees in Random Forest?
 - (c) What will happen if the random number features chosen for splitting nodes in a Random Forest is very large?
4. Under what circumstances we prefer stacking to boosting and bagging?
5. 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 all classifiers are uncorrelated*.
 - (a) First, let's assume our three independent classifiers all have the error rate of $e = 0.4$, calculated over 1000 instances with binary labels (500 A and 500 B).
 - (i) Build the confusion matrices for these classifiers, based on the assumptions above.
 - (ii) Using that the majority voting, what the expected error rate of the voting ensemble?
 - (b) Now consider three classifiers, first with $e_1 = 0.1$, the second and third with $e_2 = e_3 = 0.2$.
 - (iii) Build the confusion matrices.
 - (iv) Using the majority voting, what the expected error rate of the voting ensemble?
 - (c) 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.)