

School of Computing and Information Systems
The University of Melbourne
COMP90049 Introduction to Machine Learning (Semester 1, 2023)
Sample solutions: Week 7

1. What is the difference between “model bias” and “model variance”?
 - (i). Why is a high bias, low variance classifier undesirable?
 - (ii). Why is a low bias, high variance classifier (usually) undesirable?
2. Between “Model Bias” and “Model Variance”, which one is more harmful for the performance on test set than training set? Why?
3. During the training process, your model shows significantly different performance across different training sets. (a) What can be the reason? (b) How can we solve the issue?
4. You are developing a model to detect an extremely contagious disease. Your data consists of 4000 patients, out of which 100 are diagnosed with this illness. You achieve 96% classification accuracy.
 - (i). Can you trust the outcome of your model? Explain why.
 - (ii). What type of error is most important in this task?
 - (iii). Name at least one appropriate evaluation metric that you would choose to evaluate your model.
5. How does the choice of the `max_depth` hyperparameter for the stopping criterion affect the performance of decision trees?
6. How does the k value in k-NN algorithm affect the decision boundary between classes?
7. Explain the difference between “evaluation bias” and “model bias”.