Final Exam Practice Questions

The final exam is cumulative. You should go back through the exam 1 and 2 practice and actual exam questions.

Your actual exam will have questions similar to those below with space below each question to give answers. I’ve tried to emulate that here but I haven’t taken the time to print the questions out and hand write the answers, which means there may be too much or too little space in this document. I’ll make sure an appropriate amount of space is given on the actual exam (which can also help you gauge how detailed of an answer to give!)

* Please note that the example questions below are not exhaustive!
* You may notice there are no programming questions below (that is, no R or python syntax at all)
* There are some pseudo code questions. Here you are writing out the logic of the process and how you would go about doing it within a programming language without worrying about the syntax or the process.
* There is very little calculation required for these questions. For most answers that involve output and reading/using it, you do not need to simplify calculations.
* If you have any other questions about the content or structure of the exam, please post to the discussion forum!

1. What is the standard way for a classification tree to model the response over a particular region? That is, given a particular leaf of a classification tree, how would we make a classification over that region?
2. What issue can occur if we grow a very large regression or classification tree? If we’ve grown a large tree, how can we try to deal with this issue?
3. We discussed the idea of variable importance in (single) CART models and ensemble CART models such as the random forest model. Usually, how do we determine the variable importance?
4. When fitting a classification tree, we discussed three major node impurity measures. Name these three! (One turned out not to be great for actually determining the splits.)
5. True or False questions:
   1. CART models generally require you to standardize your predictors
   2. Boosted tree models generally require you to standardize your predictors
   3. Regularized MLR and logistic regression models generally require you to standardize your predictors
   4. The number of trees we use in a boosted model is important because we can overfit with too many trees.
   5. In deep learning models we still need to be worried about overfitting.
6. Describe what (non-parametric) bootstrapping is and how it is used to fit a random forest model for a regression task. Then describe how out-of-bag observations can be used as a prediction set.
7. What difference does stochastic gradient boosted trees do compared to a standard boosted tree model?
8. Name three advantages the XGBoost models provide over basic boosted models.
9. We discussed using Bayesian Additive Regression Trees. What are the three common ‘perturbations’ that are used in these models?.
10. Suppose we are doing a classification problem (two classes) and have two continuous predictors. What is meant by the term separable data here?
11. When using the maximal margin classifier, how do we try to determine the ‘optimal’ separating hyperplane? What is a disadvantage of selecting the hyperplane with this method?
12. When we don’t have separable data in a classification problem, we said we could use the support vector classifier. In this setting we described ‘slack variables’ associated with each observation. These variables took on 0, a value between 0 and 1, or a value greater than 1. Describe the situation where each of these values occur.
13. When we wish to apply the SVM model to a classification task with more than two levels, we discussed the one-versus-all approach. Describe how this SVM model works.
14. In SVM models, what is meant by a kernel function?
15. When doing unsupervised learning we discussed the KMeans algorithm for clustering. Describe the objective function this algorithm attempts to minimize for a given number of clusters, k. You don’t have to write out the math here if you prefer to use words.
16. When doing hierarchical clustering, what is meant by the term agglomeration?
17. When doing hierarchical clustering, how does the ‘complete’ linkage create a dissimilarity measure?
18. When doing hierarchical clustering, how does the ‘average’ linkage create a dissimilarity measure?
19. What is our major goal when using the Principal Components Analysis technique and how does PCA accomplish this goal?
20. What task is a convolutional neural network well-suited for?
21. In a convolutional neural network, what is a filter layer and what is the idea of a pooling layer?
22. In a Recurrent neural network, what is meant by a bidirectional RNN?