Practice Exam 2 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.
* I have not attempted to make the practice exam questions the same length as your actual exam (there will be fewer questions on your actual exam!)
* If you have any other questions about the content or structure of the exam, please post to the discussion forum!

1. We discussed using K-Nearest Neighbors (KNN) for classifying a response. Suppose we’ve determined an appropriate to use for a classification problem with three classes. We have three predictors, , , and . We want to use the KNN model to classify a response for a particular choice of , , and . Describe how the KNN model would go about classifying this observation.

   
   
   
   
 

1. Select true or false for each classification method.
   1. KNN provides a discriminant for classifying our observations
   2. The Naive Bayes provides a discriminant for classifying our observations
   3. Logistic Regression can include interaction terms.
   4. Shrinkage or penalized methods can be used with logistic regression
2. How does the Bayes’ classifier classify an observation? Based on your answer, why can we never apply this classifier in a real data situation.

   
   
   
   
 

1. Which two classification models we investigated directly modeled the conditional distributions ()?

1. Which three classification models we investigated were ‘generative’ models that modeled the distribution of and used Bayes’ theorem to obtain a classifier?

1. Does the bias-variance trade-off idea exist in the classification setting? Explain.

   
   
   
   

1. What is a confusion matrix? How is a confusion matrix more useful than simply looking at accuracy and misclassification rate?

   
   
   
   
 

1. Suppose we have a binary response variable. What major issue would come from treating the response as a numeric variable with values 0 and 1 and using a multiple linear regression model? Hint: Think about what our regression model is actually modeling, on average.

   
   
   
   
 

1. Consider the (binary) logistic regression model with a single predictor variable . How do we interpret the slope coefficient associated with ?

 

1. Consider the (binary) logistic regression model with a single predictor variable . How do we interpret the intercept coefficient?

 

1. Consider the (binary) logistic regression model with a single predictor variable . How would we use the model to classify an observation with ?

   
   
   
 

1. Suppose we have a categorical response with categories and a single predictor variable . When fitting a QDA model, we use the Normal distribution. What quantities do we model with a Normal distribution? Are those normal distributions related in anyway?

   
 

1. When doing LDA or QDA we end up with discriminant functions. How do these discriminant functions relate to the terms *linear discriminant analysis* and *quadratic discriminant analysis*? Note: I’m not looking for you to recall these functions. I just want you to note how they relate to the names.

   
   
   
 

1. For the Naive Bayes’ classifier we discussed the idea of using a *kernel density* estimator to model the distribution of a quantitative predictor given our response. That is, to approximate the distribution of each . Roughly describe what is meant by a kernel density estimator.

 

1. We discussed the use of generalized additive models (GAMs) for a regression task. Here we model

where each is of the form

and is the number of basis functions used to model . This is an *additive* model. Why do we call this an additive model?

   
 

1. Consider the piecewise polynomial regression model. Here we define our knots to be , …, and use the indicator functions

* in our regression equation given by

Suppose we have observations and we fit the model.

1. What is the estimate of in this model?

   
 

1. What is the estimate of in the model?

   
 

1. Suppose we fit a piecewise linear model to a data set and then also fit a linear spline to the data set. What is the big difference between these two types of models?

   
 

1. What are two methods for choosing the number and placement of the knots in a piecewise polynomial or spline model?

   
 

1. When discussing smoothing splines where we minimize a penalized version of the residual sum of squares,

* we discussed that the optimal model here was a natural cubic spline with knots at the unique values of the ’s. That seems to imply a huge number of parameters! However, we discussed the idea of effective degrees of freedom. Conceptually, what is meant by effective degrees of freedom and what ‘controls’ this quantity?

   
   
 

1. Suppose we have data on whether or not someone has heart disease (No = 0, Yes = 1) and a number of predictor Age (quantitative). We fit an QDA model. Relevant output from the process is given below.

| HeartDiseaseFac | count |
| --- | --- |
| 0 | 390 |
| 1 | 356 |

| HeartDiseaseFac | Age\_mean | Age\_sd |
| --- | --- | --- |
| 0 | 50.2 | 9.3 |
| 1 | 55.9 | 8.8 |

1. Specify the distribution we would use for Age given someone does not have heart disease.

   
 

1. Specify the distribution we would use for Age given someone does have heart disease.

   
 

1. Give the form of the posterior probabilities of heart disease and no heart disease using the QDA model for an arbitrary Age value. No need to simplify.