Q1 Cleaning the Data

10 points

1. +1.0
Removal of ID column

2a. +5.0
Correct conversion of factors

2b. +2.0
Partially correct conversion of factors (where a factor has fewer than k < 10 levels represented, say 1 -- 8 and 10, simply applying as.numeric will convert to a 1 -- k scale)

3. +4.0
Correct removal of rows with missing data

Q2 Exploratory Data Analysis

20 points

1a. +5.0
Clear and thorough exploration of the relationships between predictor variables

1b. +2.0
Some comments on the relationships between predictor variables

2a. +6.0
Clear and thorough exploration of the relationships between the predictor variables and the response variable

2b. +3.0
Reasonable explanation of relationships between predictor variables and response variable

3. +4.0
Sensible adjustment for overplotting in scatterplot matrix or equivalent clear graphical summary of relationship between predictors and response

4a. +5.0
Excellent supplementary data summaries additional to the scatterplot matrix

4b. +2.0
Good supplementary data summaries additional to the scatterplot matrix

Q3 Modelling

35 points

1. +3.0
Succinct explanation of logistic regression provided

2. +3.0
Succinct explanation of best subset selection (or why we want to reduce the number of predictors) provided

3. +1.0
Presentation of coefficients of final fitted logistic regression model

4. +2.0
Sensible commentary on coefficients of a fitted logistic regression model

5. +3.0
Sensible commentary on statistical significance of coefficients in one of the fitted logistic regression models

6a. +6.0
Correct calculation of AIC, BIC and an estimate of test error using k-fold cross-validation for the logistic regression models M_0, ..., M_p.

6b. +4.0
Correct estimation of test error using k-fold cross-validation error for the logistic regression models M_0, ..., M_p but no calculation of AIC and BIC.

6c. +2.0
Partially correct estimation of test error using k-fold cross-validation error for the logistic regression models M_0, ..., M_p but no calculation of AIC and BIC.

6d. +2.0
Correct calculation of AIC, BIC but no estimate of test error using k-fold cross-validation error for the logistic regression models M_0, ..., M_p.

7a. +4.0
Thorough and sensible justification of the choice of k in logistic regression.

7b. +2.0
Partial justification of the choice of k in logistic regression.

7c. +1.0
Choice of k without proper justification in logistic regression

8. +3.0
Succinct explanation of discriminant analysis / LDA / QDA as appropriate.

9a. +3.0
Careful justification of choice between LDA and QDA or reason for considering both

9b. +1.0
Weak justification of choice between LDA and QDA or reason for considering both

10. +1.0
Presentation of group means in LDA / QDA

11. +2.0
Sensible comment on group means in LDA / QDA

12a. +4.0
Careful rationale for choice of predictors in LDA / QDA

12b. +2.0
Partial justification for choice of predictors in LDA / QDA

Q4 Model Comparison

15 points

1. +3.0
Succint explanation of k-fold cross-validation as a method for estimating test error

2a. +2.0
Correct calculation of test error using chosen logistic regression model

2b. +1.0
Calculation of test error using chosen logistic regression model based on the validation set approach (rather than k-fold cross-validation)

3a. +2.0
Correct calculation of test error using chosen Bayes classifier for LDA / QDA

3b. +1.0
Calculation of test error using chosen using chosen Bayes classifier for LDA / QDA based on the validation set approach (rather than k-fold cross-validation)

4. +2.0
Comment on how the comparison was kept fair, e.g. by using the same folds.

5a. +6.0
Outstanding justification for choice of final model

5b. +4.0
Good justification for choice of final model

5c. +2.0
Weak justification for choice of final model

Q5 Report Writing and Presentation

20 points

1a. +4.0
Outstanding presentation

1b. +2.0
Good presentation

2a. +4.0
Outstanding structuring

2b. +2.0
Good structuring

3a. +4.0
Use of numbering and captions in Figures and Tables

3b. +2.0
Some use of captions and numbering for Figures and Tables

4a. +4.0
Outstanding scientific writing

4b. +2.0
Good scientific writing

5a. +4.0
Inclusion of well-structured Appendix

5b. +2.0
Inclusion of Appendix that lacked structure / comments