CSI 772
Homework 2
Bryan Adams

4.8.1

$$p(x) = \frac{e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}} \tag{1a}$$

$$\frac{p(x)}{1-p(x)} = \frac{\frac{e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}}}{1-\frac{e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}}} \tag{1b}$$

$$\frac{p(x)}{1-p(x)} = \frac{\frac{e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}}}{\frac{1+e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}}-\frac{e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}}} \tag{1c}$$

$$\frac{p(x)}{1-p(x)} = \frac{\frac{e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}}}{\frac{1}{1+e^{\beta_0+\beta_1\mathbf{X}}}} \tag{1d}$$

$$\frac{p(x)}{1-p(x)} = \frac{e^{\beta_0+\beta_1\mathbf{X}}}{1+e^{\beta_0+\beta_1\mathbf{X}}} * \frac{1+e^{\beta_0+\beta_1\mathbf{X}}}{1} \tag{1e}$$

$$\frac{p(x)}{1-p(x)} = e^{\beta_0+\beta_1\mathbf{X}} \tag{1f}$$

6.6.1

(a) Best subset considers the most models with the most parameters. This will provide you a model that fits your training set the best; however, it will probably overfit the data.

(b) You do not know which method will work the best with your test data; however, the method that least overfits the training data will have the smallest test data.

(c)
i. True - Model $k+1$ will have all the parameters that are identified in the models of k
ii. True - Model, as you decrease k you are removing parameters, therefore k will be a subset of $k+1$
iii. False - The methods function different. Your method used may identify different combinations of parameters to retain.
iv. False - The methods function different. Your method used may identify different combinations of parameters to retain.
v. False - With each step k you pick the best of the combinations of the k parameters. With each increase in k you could get a completely different combination of parameters.

6.6.3

A lasso regression model is fit by minimizing the provided equation.

(a) iv - at $s =$ your model is a horizontal line, as you increase s your model will fit the data better and decrease your RSS, eventually you will overfit the data with a high enough s.

(b) ii - at s your model is a horizontal line, as you increase s your model will fit the data better, but since you over fit the training data, you will most likely have your RSS increase after a certain value of s do to overfitting your training data.

(c) iii - as you increase s your model becomes more flexible which means a her variance

(d) iv - based on the bias-variance trade-off, your bias will steadily decrease with a more flexible model.

(e) v - $Var(\epsilon)$ is assumed constant

8.4.3

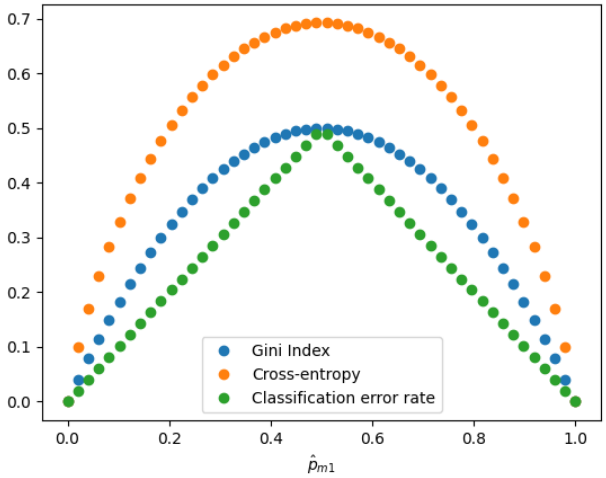

Figure 1: Plot of problem 8.4.3

8.4.5

Using **majority vote**, green has 4 and red has 6, therefore you would classify **red**

Using **average** = 0.45, you would classify **green**

9.7.2

Reference Figure 2 for *a, b, c.*

(d) Simplifying the provide equation you arrive at equation 2c, which is a linear combination of $\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_1^2}, \mathbf{X_2^2}$

$$(1 + \mathbf{X_1})^2 + (2 - \mathbf{X_2})^2 > 4 \tag{2a}$$

$$(1 + \mathbf{X_1})(1 + \mathbf{X_1}) + (2 - \mathbf{X_2})(2 - \mathbf{X_2}) > 4 \tag{2b}$$

$$1 + 2\mathbf{X_1} + \mathbf{X_1}^2 + 4 - 4\mathbf{X_2} + \mathbf{X_2}^2 > 4 \tag{2c}$$

Table 1: Classification of provide points, **c)**

Point	Classification
$(0, 0)$	Blue
$(-1, 1)$	Red
$(2, 2)$	Blue
$(3, 8)$	Blue

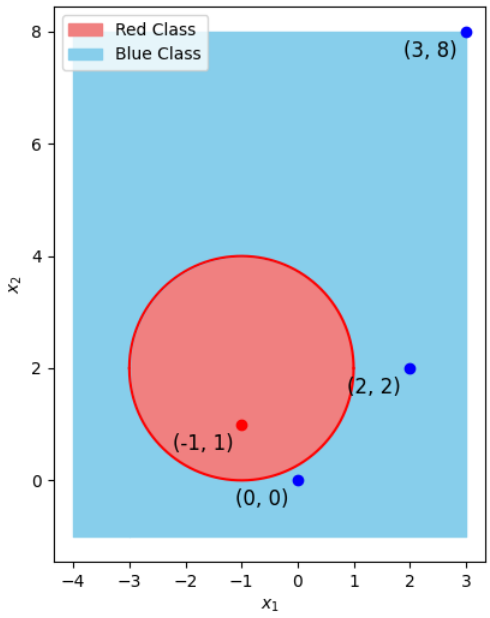

Figure 2: Plot of problem 9.7.2

11.9.1

In order for a censoring mechanism to be independent the event time T has to be independent of the censoring time C

(a) Independent - A person's phone number is not related to the event of drug relapse.

(b) Not independent - The censoring is occurring at a particular age, which is related to a person's longevity.

(c) Not independent - Sick patients will most likely die earlier, which is related to their longevity.

(d) Not independent - Since people becoming employed are not responding it would impact the measurement of unemployment duration.

(e) Not independent - Since women delivering earlier are having shorter pregnancy lengths it would impact the measurement of pregnancy duration.

(f) Not independent - Participants being censored have more years of education, which will impact the number of years of education being measured.

(g) Independent - All participants being censored occurs at the end of the study, which is independent fo the measurement.

(h) Independent - Since there is no difference between the quality of the plants, the plant, and censor time, would not be related to failure time.

(i) Not independent - Since the one plant produces better parts, the earlier one, the censor time will be related to measurement of failure time.