STAT 432 Logistic Regression

Spring 2018 / Dalpiaz / UIUC

Exercise 1

Consider a categorical response Y which takes possible values 0 and 1 as well as a single numerical predictor X. Recall that

$$p(x) = P(Y = 1 \mid X = x)$$

Consider the model

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x$$

and estimated coefficients

- $\hat{\beta}_0 = 2.4$
- $\hat{\beta}_1 = -1.2$

(a) Provide a classification when $x = 2.2$ that attempts to minimize the classification error.

(b) Calculate an estimate of $P[Y = 1 \mid X = 1]$

(c) Calculate an estimate of $P[Y = 0 \mid X = 2.5]$

(d) Find a value c that splits x into regions classified as 1 and 0. Define these regions.

Exercise 2

Consider a categorical response Y which takes possible values 0 and 1 as well as two numerical predictors X_1 and X_2. Recall that

$$p(x) = P[Y = 1 \mid X = x]$$

Consider the model

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

and estimated coefficients

- $\hat{\beta}_0 = 5$
- $\hat{\beta}_1 = -4$
- $\hat{\beta}_2 = -2$

(a) Derive and sketch the decision boundary for the classifier that results from this model. Shade the region that will be classified as **one**.

(b) Suppose the data used to estimate the coefficients has the response, y, flipped. That is, *one* becomes *zero*, and *zero* becomes *one*. What effect would this have on the:

- Decision Boundary?
- Classifications?
- Estimated coefficients?