MATH3021 Assignment 3, due on Oct 5

(1) Write down the explicit form of the likelihood function in logistic regression for two predictors, namely, as a function of β_0, β_1, and β_2.

(2) Suppose we collect data for a group of students in a statistics class with variables $X_1 =$ hours studied, $X_2 =$ undergrad GPA, and $Y =$ receive an A. We fit a logistic regression and produce estimated coefficient, $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 1$.

a). Estimate the probability that a student who studies for 40 hours and has an undergrad GPA of 3.5 gets an A in the class.

b). How many hours would the student in a) need to study to have a 50% chance of getting an A in the class?

(3) Suppose that you wish to classify an observation $X \in R$ into *apples* and *oranges*. You fit a logistic regression model and find that

$$\hat{P}(Y = orange|X = x) = \frac{\exp\{\hat{\beta}_0 + \hat{\beta}_1 x\}}{1 + \exp\{\hat{\beta}_0 + \hat{\beta}_1 x\}}.$$

Your friend fits a logistic regression model to the same data using the $softmax$ formulation and finds that

$$\hat{P}(Y = orange|X = x) = \frac{\exp\{\hat{\beta}_{orange,0} + \hat{\beta}_{orange,1} x\}}{\exp\{\hat{\beta}_{orange,0} + \hat{\beta}_{orange,1} x\} + \exp\{\hat{\beta}_{apple,0} + \hat{\beta}_{apple,1} x\}}.$$

a). What is the log odds of *orange* versus *apple* in your model?

b). What is the log odds of *orange* versus *apple* in your friend's model?

c). Suppose that in your model, $\hat{\beta}_0 = 2$ and $\hat{\beta}_1 = -1$. What are the coefficient estimates in your friend's model? Be as specific as possible.

d). Now suppose that you and your friend fit the same two models on a different data set. This time, your friend gets the coefficient estimate $\hat{\beta}_{orange,0} = 1.2$, $\hat{\beta}_{orange,1} = -2$, $\hat{\beta}_{apple,0} = 3$, $\hat{\beta}_{apple,1} = 0.6$. What are the coefficient estimates in your model?

e). Finally, suppose you apply both models from d) to a data set with 2000 test observations. What fraction of the time do you expect the predicted class labels from your model to agree with those from your friend's model? Explain your answer.

(4) Use the logistic regression and linear discriminant analysis to predict the direction of stock price. You may use 80% of samples as a training set and 20% of samples as the test set.