Statistical and Machine Learning
ECON576
Problem Set 3

1. (10 points) This question involves the `Weekly` dataset, which is included in the package `ISLR2`.

   (a) Fit a logistic regression with `Direction` as the response and the five lag variables plus `Volume` as predictors. Use the `summary()` to print the results.

   (b) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$?

   (c) Compute the confusion matrix and overall fraction of correct predictions.

   (d) Repeat (c) using LDA. Which method performs better on this data?

2. (10 points) Consider the logistic regression with one predictor such that the success probability is given by

$$p(x_i; \boldsymbol{\beta}) = \mathbb{P}(y_i = 1 \,|\, x_i, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)},$$

   where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Recall that the likelihood function for this model is

$$L(\boldsymbol{\beta} \,|\, \mathbf{y}) = \prod_{i=1}^{n} p(x_i; \boldsymbol{\beta})^{y_i} (1 - p(x_i; \boldsymbol{\beta}))^{1-y_i}.$$

   (a) Show that the log-likelihood function can be written as

$$\ell(\boldsymbol{\beta} \,|\, \mathbf{y}) = -\sum_{i=1}^{n} \log(1 + \exp(\beta_0 + \beta_1 x_i)) + \sum_{i=1}^{n} y_i (\beta_0 + \beta_1 x_i).$$

   (b) Compute the gradient of the log-likelihood, $\nabla \ell(\boldsymbol{\beta} \,|\, \mathbf{y}) = \left(\frac{\partial \ell}{\partial \beta_0}, \frac{\partial \ell}{\partial \beta_1} \right)'$.

   (c) The dataset `logit_data.csv` contains 500 observations of a student's grade and her average weekly hours of study. More specifically, the first column contains a binary variable of whether the student gets an A from a class ($y = 1$) or not ($y = 0$); the second column contains the student's average weekly hours of study (x). Use the built-in function `glm` to obtain the maximum likelihood estimate of $\boldsymbol{\beta}$.

   (d) Use gradient ascent to find the maximum likelihood estimate of $\boldsymbol{\beta}$. That is, given a positive learning rate $\alpha > 0$ and an initial value $\boldsymbol{\beta}_0$, iteratively update $\boldsymbol{\beta}_{t+1}$ using:

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + \alpha \nabla \ell(\boldsymbol{\beta} \,|\, \mathbf{y}) \big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_t}.$$

   The iteration stops until $||\boldsymbol{\beta}_{t+1} - \boldsymbol{\beta}_t|| < \varepsilon$, where $||\cdot||$ is the usual Euclidean norm, i.e., $||\mathbf{z}|| = \sqrt{z_1^2 + z_2^2}$. Use the initial value $\boldsymbol{\beta}_0 = (0, 0)'$, $\alpha = 0.001$ and $\varepsilon = 10^{-8}$. What are the maximum likelihood estimates of β_0 and β_1? Report the values of the intermediate $\boldsymbol{\beta}_t$ using a scatter plot.

(e) If a student studies for $x = 10$ a week on average, what is the probability that she will not get an A (i.e., $y = 0$)?