Department of Electrical and Computer Engineering
University of Delaware
FSAN/ELEG815 Analytics I: Statistical Learning
Homework #6, Fall 2019

Name: ____

1. Consider a Wiener filter process characterized by the correlation matrix $\mathbf{R}$ of the tap input vector $\mathbf{u}(n)$. $\mathbf{R}$ is given by $\mathbf{R} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ and the cross-correlation vector of $\mathbf{u}(n)$ and the desired response $d(n)$ is $\mathbf{p} = \begin{bmatrix} 0.5 & 0.25 \end{bmatrix}^T$

   (a) Evaluate the tap weights of the Wiener filter.

   (b) Find the minimum mean-square error for this filter.

   (c) Formulate a representation of the Wiener filter in terms of the eigenvalues and eigenvectors of $\mathbf{R}$.

2. Revisit the handwritten digit recognition problem (reproduce results from of chapter "Nonlinear Transformation"). Separate digit 1 from all the other digits, using intensity and symmetry as your inputs variables like you did before. Use the same data set sent to you before,

   - Use linear regression for classification. Even though, linear regression learns a real-valued function, binary-valued functions are also real-valued $\pm 1 \in \mathbb{R}$. Thus, you can use linear regression to compute $\mathbf{w}$ and approximate your binary classification $\mathbf{w}^T \mathbf{x}_n \approx y_n = \pm 1$. Use your result for $\mathbf{w}$ to compute $\text{sign}(\mathbf{w}^T \mathbf{x}_n)$ and report the value for E_{in} and E_{out}.
   - Repeat item (a) with a third-order polynomial transform Φ_3 to get a different representation of the data.
   - Compare E_{in} and E_{out} from (a) and (b).
   - Show same plots of Slide 14 of chapter "Nonlinear Transformation and Logistic Regression'.

   Note: You are not allowed to use prebuilt functions (implement linear regression algorithm).

3. Examine some numerical and graphical summaries of the "Weekly" data. This data set consists of 1,089 weekly returns for 21 years, from the beginning of 1990 until the end of 2010. For each week, we have recorded the percentage returns for each of the five previous trading weeks, **Lag1** through **Lag5**. We have also recorded **Volume** (the number of shares

traded on the previous week, in billions), **Today** (the percentage return on the week in question) and **Direction** (whether the market was Up or Down on this week).

(a) Compute the correlation matrix that contains all of the pairwise correlations among the predictors in the data set. Plot **Volume** over time. Draw your conclusions. Do there appear to be any patterns? (Note: exclude the output **Direction**).

(b) Use the full data set to perform a logistic regression with **Direction** as the response and the five **lag** variables plus **Volume** as predictors. Report the coefficients ($\mathbf{w}$). Draw your conclusions. Do any of the predictors appear to be significant?

(c) Report the probability output $g(\mathbf{x})$ for the first 10 data points of your training set.

(d) Compute the overall fraction of correct predictions (using the training set). Hint: Use 0.5 as a threshold for the probability $g(\mathbf{x})$ to obtain a binary output.

(e) Now fit the logistic regression model using a training data period from 1990 to 2008, with **Lag2** as the only predictor. Compute the overall fraction of correct predictions for the held out data (that is, the data from 2009 and 2010).

Note: You are not allowed to use prebuilt functions (implement Logistic regression algorithm).

4. For logistic regression, show that

$$\begin{aligned}\nabla E_{in}(\mathbf{w}) &= -\frac{1}{N}\sum_{n=1}^{N}\frac{y_n\mathbf{x}_n}{1+e^{y_n\mathbf{w}^T\mathbf{x}_n}} \\ &= \frac{1}{N}\sum_{n=1}^{N}-y_n\mathbf{x}_n\theta(-y_n\mathbf{w}^T\mathbf{x}_n)\end{aligned}$$

Argue that a 'misclassified' example contributes more to the gradient than a correctly classified one.

5. Let $x_1, x_2, ..., x_n$ be i.i.d with Poisson distribution $P(\lambda)$:

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!},\ k = 0, 1, 2, ..., \infty \tag{1}$$

a. Find the maximum likelihood estimate of λ.