Department of Electrical and Computer Engineering
University of Delaware
FSAN/ELEG815 Analytics I: Statistical Learning
Homework #5, Fall 2019

Name: ____

1. Handwritten Digit Recognition. The goal is to recognize the digit in each image of the dataset given in "DigitsTraining" which contains some digits from the US Postal Service Zip Code. We are going to decompose the big task of separating ten digits into smaller tasks of separating two of the digits (binary classification). Use two digits: the final number in your UD ID and conveniently choose any other number to replicate the results from the slides chapter "The Learning Problem" (take into account the features that you are going to use for classification to choose the second number).

   (a) Extract 2 features from the images: average intensity and symmetry. Using this two features, implement the Perceptron Learning Algorithm. Use an error metric for binary classification. To compute E_{out}, use the testing set given to you in "DigitsTesting". Show only 200 iterations.

   (b) Repeat item (a), for the pocket algorithm. Show the same two plots that are in Slide 29 and 30 of the chapter "The Learning Problem". Compare your results and draw conclusions.

   (c) 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}.

   (d) Repeat item (b), using your result for $\mathbf{w}$ in item (c), as the initial weights in the pocket algorithm. Compare your results.

   (e) Extract one more feature from the images that could help to improve your previous results. Describe how you compute this feature and why is it representative of your data?.

   (f) Repeat items (a), (b), (c), and (d), using the three features (average intensity, symmetry and the one that you choose in (e)).

   Dataset description: The first column in DigitsTraining and DigitsTesting corresponds to the digit number, following columns correspond to 256 pixels of the 16×16 pixel image of the digit. Thus, we have 7291 inputs in DigitsTraining and 2007 inputs in DigitsTesting. From these datasets, work only with those inputs that correspond to the digits you chose. Remember, one of the digits corresponds to the final number in your UD ID.

2. The linear regression weight vector $\mathbf{w}_{lin}$ produces an estimate of $\mathbf{y}$ i.e. $\hat{\mathbf{y}} = \mathbf{X}\mathbf{w}_{lin} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. The estimate $\hat{\mathbf{y}}$ is a linear transformation of the actual $\mathbf{y}$. Consider the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, where $\mathbf{X}$ is an N by $d+1$ matrix, and $\mathbf{X}^T\mathbf{X}$ is invertible.

   (a) Show that $\mathbf{H}$ is symmetric.

   (b) Show that $\mathbf{H}^K = \mathbf{H}$ for any positive integer K.

   (c) If $\mathbf{I}$ is the identity matrix of size N, show that $(\mathbf{I} - \mathbf{H})^K = \mathbf{I} - \mathbf{H}$ for any positive integer K.

   (d) Show that $\text{trace}(\mathbf{H}) = d + 1$.

3. Remember the inequality for multiple hypotheses:

$$\mathbb{P}[|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2Me^{-2\epsilon N}$$

   If we replace M by $m_{\mathcal{H}}(N)$ which can be bounded by a polynomial, the generalization error will go to zero as $N \to \infty$ which implies learning is feasible. To prove this, assume $m_{\mathcal{H}}(N)$ can be bounded by the polynomial N^{k-1} and compute the following simplified limit for $\epsilon > 0$ and k being a finite positive integer (i.e. $k > 0$ and $k < \infty$):

$$lim_{N\to\infty} \;\; N^{k-1}e^{-\epsilon N} \tag{1}$$