ECE 5403: Machine Learning **(Due: 04/06/2020)**

Submission Assignment #2

Instructor: Chun-Shu Wei *Name:* Student name, *Student Id:* Student Id

Course Policy: Read all the instructions below carefully before you start working on the assignment, and before you make a submission. For this assignment, please hand in the following two things: a pdf file and a ipynb file.

- PDF file: contains both your results and explanations. Please name this pdf file as **HW2_StudentID_Name.pdf** and **remember to type your Student ID and Name in pdf (e.g. HW2_9400000_chunshuwei)**.
- Ipynb file: write the comment to explain your code. Please name this ipynb file as **HW2_StudentID_Name.ipynb**
- Please name your assignment as **HW2_StudentID_Name.zip**. The archive file contains source code(ipynb file) and report (pdf file).
- Implementation will be graded by completeness, algorithm correctness, model description, and discussion.
- PLAGIARISM IS STRICTLY PROHIBITED.
- Please submit your assignment as ONE single zip file on the E3 system. Paper submission is not allowed. Inserting clear scanned image of handwritten derivations is accepted. Denote date and time on the first page.
- Submission deadline: **2020.04.06 11:59:59 PM**.

Problem 1: Python code Exercise (12.5+12.5+12.5+12.5=50 points)

Data: MEAP93.csv
Notice: Please hand in with your code and results.
Note: Do NOT use a ready-made regression function in python. (e.g. sklearn.linear_model.LinearRegression, statmodels)
Let **math10** denote the percentage of tenth graders at a high school receiving a passing score on a standardized mathematics exam; **lnchprg** denote the percentage of students who are eligible for the lunch program.

$$ma\hat{t}h10_i = \hat{\beta}_0 + \hat{\beta}_1 log(expend)_i + \hat{\beta}_2 lnchprg_i,\ i = 1, 2, ..., 408 \tag{0.1}$$

The loss function is l_2 loss function:

$$\|math10 - ma\hat{t}h10\|_2^2. \tag{0.2}$$

Where $math10$ and $ma\hat{t}h10$ are vectors. We can rewrite our linear regression problem as:

$$\min \|math10 - ma\hat{t}h10\|_2^2. \tag{0.3}$$

And we define :

$$
\begin{aligned}
&y_{408\times1} = math10\\
&X_{408\times3} = \begin{bmatrix} 1 & log(expand)_1 & lnchprg_1 \\ 1 & log(expand)_2 & lnchprg_2 \\ ... & ... & ... \\ 1 & log(expand)_{408} & lnchprg_{408} \end{bmatrix}\\
&F(w^k) = \|y - Xw^k\|_2^2\\
&w^{k+1} = \begin{bmatrix} \hat{\beta_0}^{k+1} \\ \hat{\beta_1}^{k+1} \\ \hat{\beta_2}^{k+1} \end{bmatrix} = w^k - \eta\nabla F(w^k),\ w^0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\\
&y\_pred^k_{408\times1} = \hat{math10}^k = Xw_k\\
&\eta = learning\ rate\\
&tol = tolerance = 1e-5\\
&sse = \|y\_pred^{optimal} - y\|_2^2\\
&error^k = \|w^k - w^{k+1}\|_2,\ error^0 = 1
\end{aligned}
\tag{0.4}
$$

where k denotes the number of iteration. The algorithm converges as $error^k < tol = 1e-5$ that provides the best estimate $y\_pred^{optimal} = Xw^{optimal}$.

(a) Please solve (0.3) by gradient descent where $\eta = 1e-10$. What is sse and k? Does this algorithm converge?

(b) Please solve (0.3) by gradient descent where $\eta = 1e-5$. What is sse and k? Does this algorithm converge?

(c) Compare your result in (a) and (b). If your result in (a) is different from result in (b), please explain the possible reason.

(d) Please calculate sse in Problem 2 (d) in HW1; compare it with sse you got in (a). What is your observation?

Problem 2: Python code Exercise (12.5+12.5+12.5+12.5=50 points)

Load the training set and testing set from the following website link: http://yann.lecun.com/exdb/mnist/. It contains a handwritten digit collection of 28×28 images that is used for handwriting recognition benchmarks. Please use the training set to train your classifier, and use the testing set for evaluation.

Note: DO NOT use the testing set for training.

(a) Train classifiers using nearest neighborhood(1NN) and K-nearest neighborhood(KNN).

(b) Train a classifier using the support vector machine(SVM) with linear kernel.

(c) Perform kernel density estimation(KDE) on each class, and classify the digits by finding the distribution with highest probability they belong to.

(d) Compare the classification performance, training complexity and testing complexity between these methods and explain the reasons in your .pdf file.