COSC 6330: Advanced Machine Learning, Homework2
Due 03/06/2020 at 11:30pm

Rules:

Late submission penalties:
1- **Within 48 hours from the due date, you get 50% of your score**
2- **After 48 hours from the due date, you get zero**

Please note: programming assignment instruction:
In your code, you cannot use machine learning libraries such as those available from scikit-learn for learning the models or for cross-validation.

Deliverables:
You are required to submit the following with the given file names:
- LRDigit (for logistic regression) and GNBDigit (for naiive bayes): Your code, with instructions on how to compile and run your code, and any other necessary details or files. If I can not run your code, you get zero
- HW2.pdf: A PDF file that contains a description of the algorithm and a screenshot of your output. Description must be your own writing! Do not copy and paste from web! As well as answers to the questions and problems

……………………………………………………………………………………………………

1- Comment on the convexity of the following optimization problem.

$$\underset{x,y}{\text{minimize}} \quad f(x,y) = \frac{1}{2}(\alpha x^2 + \beta y^2) - x$$
$$\text{subject to} \quad (x,y) \in \mathbb{R}^2.$$

2- Programming assignment:

The digit dataset has been uploaded in D2L is taken from sklearn. Each datapoint is an 8x8 image of a digit. There are

Classes	10
Samples per class	~180
Samples total	1797
Dimensionality	64
Features	integers 0–16

The classes are digits 0,1,2,...,9. The last column shows class label that we are trying to classify. All other columns are data values.

Using digits dataset, train and evaluate two classifiers:

(a) Logistic regression (LR)
(b) Naive-Bayes (NB) with marginal Gaussian distributions, i.e., Gaussian Naïve Bayes

-Do 80-20 train-test splits and 10-fold cross validation. Show train and test error tables for both (a) and (b).

-Compare the performance of LR vs NB.