COSC 6330: Advanced Machine Learning, Homework1
Due 02/11/2020 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:

- LDAGaussianDigit.: 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
- HW1.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- Please summarize the key points for each of the lectures. (less than 5 lines for each lecture)

a) lecture 1
b) lecture 2
c) Lecture 3
d) Lecture 4
e) Lecture 5

2- Please provide an expression for the probability density function $p(x;\mu,\Sigma)$ of a multivariate Gaussian distribution with mean μ and covariance Σ in terms of x, μ, Σ?

3- Suppose a multi-class logistic regression has been used as a model for digit recognition as shown below

$$P(Y=k|X=\mathbf{x}) = \frac{\exp(\mathbf{w}_k^T\mathbf{x})}{1+\sum_{l=1}^{K-1}\exp(\mathbf{w}_l^T\mathbf{x})} \quad \text{for} \quad k=1,\ldots,K-1$$
$$P(Y=K|X=\mathbf{x}) = \frac{1}{1+\sum_{l=1}^{K-1}\exp(\mathbf{w}_l^T\mathbf{x})},$$

a) Which parameters need to be estimated?
b) How many parameters in total need to be estimated?
c) With n training samples as:

$$\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_n,y_n)\}$$

d) write down explicitly the log-likelihood function for multi-class logistic regression (as shown below) and simplify it.

$$L(\mathbf{w}_1,\ldots,\mathbf{w}_{K-1}) = \sum_{i=1}^{n}\ln P(Y=y_i|X=\mathbf{x}_i).$$

e) Take the **gradient** with respect to w and simplify.

4- 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.

Apply Fisher's linear discriminant analysis (LDA) to the dataset (both S_B and S_W computed from the data) to project data to two dimensions. Then apply Gaussian generative modeling (bi-variate) and do 10-class classification of digits. Show training and test set error rates from 10-fold cross validation.

Hint: you need to estimate and use class priors π_k and parameters(μ_k,Σ_k), k=1,...,10.

Error rate: Error rate is the percentage of wrongly classified data points divided by the total number of classified data points.