Shiraz University

Statistical Pattern Recognition (Fall 2020)

Homework#2: Logistic, Softmax Regression (Multiclass Classification), and Bayesian Classification

Due date: 22th December 2020

In order to do this homework, you have to go through logistic, softmax regression, and Bayesian classification theories and concepts.

Part A: Logistic, Softmax Regression (Multiclass Classification)

Dataset: Iris https://archive.ics.uci.edu/ml/datasets/Iris

As mentioned in previous homework the Iris dataset consists of 4 features and 3 classes. Use all features and classes for this part of homework. In before homework you used logistic regression for binary classification. In this part you should use whole iris dataset for multiclass classification (one-vs.-one and one-vs.-all) by logistic regression. Then using softmax regression and compare them.

- Consider the first 80% of the data in **each class** for train and the rest 20% for test.
- Train multiclass classification (one-vs.-one and one-vs.-all) by logistic regression and report train and test accuracy for both of method.
- Plot cost function for enough iteration for one-vs.-all method and report convergence iteration in this method.
- Train multiclass classification by softmax regression and report train and test accuracy.
- What method (one-vs.-one, one-vs.-all or softmax) has worked best?

Part B: Bayesian Classification

Dataset: BC-Train1, BC-Test1, BC-Train2, BC-Test2

In this part you have two datasets (Dataset1= BC-Train1, BC-Test1, Dataset2 = BC-Train2, BC-Test2). Each dataset contains two classes and each class generated from one Gaussian distribution. In this part you have to construct two Bayesian classifiers so as to classify both train and test data.

- Use a Bayesian classifier to classify both the train and test datasets and calculate both accuracies (for each dataset separately).
- Plot the decision boundary and classification results while representing the misclassified samples with a different color or shape (for each dataset separately).
- Plot estimated PDFs. (3D for each dataset separately)
- Contour estimated PDFs along with the decision boundary. (2D for each dataset separately)
- What is the main difference between two datasets? Explain your answer using your results and plots.

Notes:

- **Pay extra attention to the due date. It will not extend.**
- **Be advised that submissions after the deadline would not grade.**
- **Your implementation should be functional.**
- **Prepare your full report in PDF format and include the figures and results.**
- **Do not use sklearn or any similar library for logistic, softmax regression and Bayesian classification and write your own code.**
- **The allowed programming languages are any language and feel free.**
- **Feel free for using sklearn in python for load iris or split train and test dataset.**
- **Feel free for using numpy, pandas, or any regular library in python.**
- **Submit your assignment using a zipped file with the name of "StdNum_FirstName_LastName.zip" to csehws.shirazu.ac.ir or for someone has problem with it to compuscien@gmail.com with SPR-Fall 2020-HW#1 subject.**