Problem assignment 5

Due: Thursday, February 22, 2018

In this problem we shall investigate the ”Pima” dataset and learn classification models for it. Recall we performed some exploratory analysis of the Pima dataset in Problem set 1. You can download the dataset (pima.txt) and its description (pima desc.txt) from the course web page. In addition to the complete dataset pima.txt, you have pima train.txt and pima test.txt you will need to use for training and testing purposes. The dataset has been obtained from the UC Irvine machine learning repository: <http://www1.ics.uci.edu/_mlearn/MLRepository.html>.

Problem 1. Logistic regression model

First we try the logistic regression model in combination with gradient methods. Give solutions to the following tasks:

(a) Use functions from assignment 4 to normalize the inputs in the pima dataset based on the data in the training set. Generate two new files *pima­\_train\_norm.txt* and *pima\_test\_norm.txt*.

Part1.m

(b) Familiarize yourself with a batch-mode gradient procedure in file *Log\_regression.m*, in which all data points are considered at the same time.

(c) Implement and submit a program ***main1.m*** that runs the gradient procedure on the training dataset for 2000 iteration steps (also called epochs). Initialize all weights to 1 at the beginning. Use 2/aqrt(i) learning rate schedule.

(d) Include graph functions for monitoring the progress of misclassification errors (for both training and testing data) in ***main1.m*** as used in the previous problem set (HW-5).

In the report include final:

– Training and test misclassification errors

– Confusion matrices for the train and test sets

– Sensitivity and specificity of the model on the test set.

(e) Experiment with the learning algorithmby changing initial weights, learning schedule, number of epochs. Report training and test misclassification errors. What was the best result you could get?