Math 354: Class Exercise
Machine Learning II
Instructor: Dr. Fred Park

The goal for machine learning, in this context is, for the computer to be able to learn from data without being programmed. An example would be for the computer to learn a model for pass/no-pass based on exam scores and labeled data. By labeling, we mean that a value of 1 means pass and 0 no-pass. Thus, if two midterm scores were input that were not part of the training set, the computer would be able to tell you if they passed or not without ever knowing the specific grading criteria. (In fact, it can also learn the grading criteria if needed.) An example: if scores were 40 and 80 for midterms 1 and 2 respectively, the computer would tell you that the student did not pass. However, if the scores were 40 and 90, it would tell you that they did in fact pass. Thus, the computer learned a decision boundary as to what scores would allow a pass or not. Based on this, we can make predictions on pass/no-pass of future students.

Logistic Regression amounts to maximizing the likelihood of the parameters θ and b in the following energy:

$$\ell(\theta, b) = \sum_{i=1}^{m} y^{(i)} \log h(\vec{x}^{(i)}) + (1 - y^{(i)}) \log(1 - h(\vec{x}^{(i)})). \tag{1}$$

Here $h(z) = \frac{1}{1+e^{-z}}$ and $\vec{x}^{(i)} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \end{bmatrix}$ a training example.

1. Write out the gradient ascent and stochastic gradient ascent for the Logistic Regression energy.

2. From fredpark.com/teaching download the two files:
   `scatter_plot_dec_boundary.m`
   `exam_scores_training.txt`
   Run the `scatter_plot_dec_boundary` script and you will see the plot of the exam scores and pass/no-pass values on a scatter plot. A scatter plot is simply plotting one feature against the other with labels ID'd in some way. You did this in #1 for a hypothetical problem.

   (a) Find a decision boundary and by training on the supplied training data.

   (b) If we store the scores of a testing data-set as a 1×3 vector, for example $x_1 = [1\ 40\ 80]$ and $x_2 = [1\ 40\ 90]$. Here, the 1 is needed as a convention and does not indicate any information on passing or not. 40 = MT 1 score and 80 and 90 is MT 2 score for x_1 and x_2 respectively. Can you find a way to apply your decision boundary directly to these vectors to obtain a pass or not decision?

   (c) Test on the following cases:

Student	MT#1	MT#2
1	60	65
2	61	75
3	99	22
4	22	99
5	51	89
6	40	85
7	70	62

3. From this, can you figure out what the exact grading criteria was? Is it even necessary?

4. Training error is error involving the number of misclassified data points when training. Since we are using a linear decision boundary, it is reasonable to have training error as long as it is small. Usually, it is the ratio of the misclassified data points to the total number of data points which yields a percentage. What is the training error for your classifier in this problemm ($\longleftarrow$ typo!)?

5. Test your decision boundary on the testing data supplied.

6. Testing or classification error is the ratio of the number of misclassified points to the total number of points. It is a percentage. Usually, some of the training data is withheld so that once a classifier is developed from the training set, we can test it on the testing set of data. Can you think of a way to withhold 10% of the data to this problem. Re-create your decision boundary by training on the 90% left. Then test on the testing set of 10% points withheld to obtain a testing error. What is it?