CISC 484/684
Homework 1
Due: Thursday, March 1, 2018 in class
No late submissions

1. You have to train and test three types of models: linear regression, logistic regression and perceptrons. All three models will be trained on the same training dataset. Only one attribute is used and hence the learned models will all be characterized by two numbers (weight for the one attribute and the threshold/intercept). The Readme1.txt file describes how to train and test the models.

   The training data meets the following criteria: If attribute value is ≥ 1 then output is 1 and if attribute value is ≤ -1 then output is 0. Thus the training data does not specify what the output value is when the attribute value is > -1 but < 1. Also, by including linear regression, we are trying to investigate whether a regression model can be used to classify by providing training instances whose output values are 0 or 1.

   a. First train and test the three models as described in the "readme" file. Observe the weights learned in the three cases. Next create 10 new test instances whose attribute values are strictly between -1 and 1. Observe how the models generalize on this different type of data. For your answer, provide your test instances and the predictions of the 3 models.

   b. Provide an instance with a target value which is consistent with the criteria (described above) in developing the training sets. This instance should result in different behavior (e.g., different sides of the lines) from linear regression and perceptron. This should be the case even if we output 0 or 1 as linear regression predictions depending on which side of the regression "line" the datapoint lies.

   First, provide the example you chose, what the two models predict. Then, explain why perceptron and linear regression come up with such different models despite the fact the data provided to them are the same and the update rule *appears* to be the same, i.e., $\Delta w_i = -\eta(y - t)x_i$ where y is the models output value for one instance and t is the target value for that instance.

2. Again, you will train and test a logistic regression model. This time the training data characterizes an instance by a set of attributes (see Readme2.txt) and the target value specifies whether the individual has diabetes or not.

   Before running the training and testing (so that you don't know the weights for these attributes), based on the attributes, come up with 3 instances (i.e., choices of attribute values) such that you believe that the indovidual will be diabetic, 3 more instances where you believe that the indovidual will *not* be diabetic and finally 4 instances where you are not

sure (since some attribute values might indicate diabetes=1 and some might indicate the opposite.

Then train model and see what it assigns to these 10 cases (by modifying the test file (see Readme2.txt). Your answer should include (i) the 10 instances you chose, (ii) whether or not you think the individual will be predicted to have diabetes or not and (iii) the model's prediction. Also show the weights it assigns to each attribute.

3. This question concerns training a perceptron. We discussed in class why a perceptron can't learn the exclusive-OR function. Now consider mapping an instance $\langle x_1, x_2 \rangle$ to another instance in a 3 dimensional space given by $\langle x_1, x_2, x_1.x_2 \rangle$. Provide a perceptron, ie., provide weights (don't train any model but manually figure out the weights) for the three dimensions and the threshold so that the 4 instances in exclusive-OR are given the appropriate outputs by your perceptron.

4. Suppose a logistic regression model has weights (including intercept) given by $\langle w_0, w_1, \ldots, w_n \rangle$. Recall that the final predictions of the logistic regression model is either 1 or 0 and that the prediction is dependent on whether the probability estimated by the model is at least 0.5. Further, Recall that this probability is given by using the model's parameter and the sigmoid function.
   Is the logistic regression model a linear separator or not? Write an expression involving the model's parameters, $w_0, w_1, \ldots, w_n$ and indicate what the separating boundary look like.