Mehryar Mohri
Introduction to Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 2
October 10, 2011
Due: October 24, 2011

Perceptron Algorithm

Download the following data sets from the UC Irvine ML repository:

`http://archive.ics.uci.edu/ml/datasets/Iris`
`http://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+(Sonar,+Mines+vs.+Rocks)`
`http://archive.ics.uci.edu/ml/datasets/Spambase`

The first two are linearly separable, the third one is not. For each data set, reserve the first half for training, the second half for testing.

1. Implement the Perceptron algorithm or use the Weka software library instead. Run it on the two separable data sets. How many updates were made by the algorithm? Compare with the upper bound known for the perceptron algorithm. What is the margin ρ_0 of the solution obtained?

2. As we saw in the midterm exam, when the training sample S is linearly separable with a maximum margin $\rho > 0$, there exists a modified version of the Perceptron algorithm that returns a solution with margin at least $\rho/2$ when run cyclically over S. Furthermore, that algorithm is guaranteed to converge after at most $16R^2/\rho^2$ updates, where R is the radius of the sphere containing the sample points.

   Fix $\epsilon \in (1/2, 1)$. Generalize that algorithm to guarantee that under the same conditions the solution has margin at least $\rho(1-\epsilon)$ (give the pseudocode). Adapting the proof given for the mid-term questions, show that the number of updates is upper bounded by $\frac{R^2/\rho^2}{(1-\epsilon)(\epsilon-1/2)}$.

   More generally, it can be proven that the same algorithm achieves a margin of at least $\rho(1-\epsilon)$ for any $\epsilon \in (0,1)$ with at most a polynomial number of updates in $O\big(\frac{R^2/\rho^2}{(1-\epsilon)}\big)$.

3. Implement the algorithm of the previous question (or modify Weka's perceptron code). Use $\rho = \rho_0$ and run the algorithm with $\epsilon = 1/4$ on the same

datas set as the first question. How many updates are made by the algorithm? What is the margin of the solution obtained? Compare with ρ_0.

4. Run the perceptron algorithm on the third data set and stop it after n passes over the training data with $n = 10, 50, 100$. Report the test error of the solution obtained in each case.