E0 270 Machine Learning *Due:* Apr 17, 2012

Assignment 6

TAs: Raman Sankaran (ramans@csa), Arun Rajkumar (arun_r@csa)

Submission instructions: Assignment solutions must be prepared in LaTeX and the PDF file together with any code must be submitted **both electronically *and* as a hard copy in class, before the start of class on the date above**. Electronic submission can be done by clicking on the following link. Any plots/tables/figures must be included in your PDF file/hard copy. Questions about the assignment can be directed to the TAs above. Extra credit problems are optional.

---

1. **Warm-up.** Let $S = \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\|_2 = 1\}$. Let $\mathbf{u}, \mathbf{w} \in S$. Compute the probability

$$\mathbf{P}_{\mathbf{x} \in S}\Big[\mathrm{sign}(\mathbf{w} \cdot \mathbf{x}) \neq \mathrm{sign}(\mathbf{u} \cdot \mathbf{x})\Big],$$

where $\mathbf{P}_{\mathbf{x} \in S}$ denotes probability over the random draw of $\mathbf{x}$ from the uniform distribution over S. Explain your answer.

2. **Perceptron.** You are given two data streams, `data_stream_1.mat` and `data_stream_2.mat`, each containing 1000 instances $\mathbf{x}_t \in \mathbb{R}^2$ and corresponding labels $y_t \in \{\pm 1\}$, $t \in \{1, \ldots, 1000\}$; you are also given two MATLAB routines `get_instance(`s`,`t`)` and `get_label(`s`,`t`)`, which return the t-th instance in data stream s and the t-th label in data stream s, respectively ($s \in \{1, 2\}$). The instances in `data_stream_1.mat` happen to be drawn iid from the uniform distribution over the surface (perimeter) of the unit ball (circle) in $\mathbb{R}^2$, i.e. from the uniform distribution over the set $S = \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\|_2 = 1\}$. The instances in `data_stream_2.mat` are generated as follows: $\mathbf{x}_1$ is set to be $(1, 0)$; then for each $t \geq 2$, $\mathbf{z}_t$ is drawn uniformly at random from $[-1, 1]^2$, and $\mathbf{x}_t$ is set to $(\mathbf{x}_{t-1} + \mathbf{z}_t)/\|\mathbf{x}_{t-1} + \mathbf{z}_t\|_2$. Thus the instances in `data_stream_2.mat` are not independent. In both cases, the labels y_t are given by a fixed linear separator, $y_t = \mathrm{sign}(\mathbf{u} \cdot \mathbf{x}_t)$, where $\mathbf{u} = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$.

   (a) Write a piece of MATLAB code `perceptron.m` which implements the perceptron algorithm on a given data stream, where instances and labels are obtained via the `get_instance(`s`,`t`)` and `get_label(`s`,`t`)` routines. Apply the algorithm to each of the two data streams above, and in each case, plot the ratio of the number of mistakes made relative to the total number of examples seen as a function of the number of examples seen. In each case, also plot an upper bound on this ratio obtained from the mistake bound discussed in class.

   (b) In the case of `data_stream_1.mat`, where the instances are drawn iid from a fixed distribution, it can also be of interest to evaluate the performance of learned models in terms of expected error on new instances drawn from the same distribution. Write a small MATLAB function `error_uniform.m` which takes as input two vectors $\mathbf{u}, \mathbf{w} \in \mathbb{R}^2$, and implements your solution to Problem 1 to compute the expected misclassification error on a new example $\mathbf{x}$ drawn uniformly from S, when the true label is $\mathrm{sign}(\mathbf{u} \cdot \mathbf{x})$ and the predicted label $\mathrm{sign}(\mathbf{w} \cdot \mathbf{x})$. Use this MATLAB function to compute the expected error of the linear predictor learned by the perceptron algorithm on each iteration, and plot this error as a function of the number of examples seen.

3. **Perceptron-based active learning.** Apply the perceptron-based active learning algorithm described in class to the above two data streams and contrast the results with those obtained in Problem 2. Specifically, write a piece of MATLAB code `active_perceptron.m` that implements the perceptron-based active learning algorithm, and apply it to both data streams. In each case, plot the number of labels requested as a function of the number of examples seen, as well as the ratio of the number of mistakes made relative to the total number of examples seen as a function of the number of examples seen. Compare these plots with those obtained in Problem 2(a) above. In the case of `data_stream_1.mat`,

also plot the expected error on a new example (drawn from the uniform distribution over S as before) as a function of the number of *labels* requested. Compare this with the plot obtained in Problem 2(b) above. What do you observe?

4. **Extra credit.** How would you extend the solution to Problem 1 above to n dimensions, where S is now the surface of the unit sphere in n dimensions, $S = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 = 1\}$?