ECE 2372 - Homework 3

Problem 1: [Convergence of PLA] Prove that the perceptron learning algorithm will converge (eventually) to a linear separator for linearly separable data set. For mathematical simplicity assume that the starting point is $\theta^0 = 0$ and for iterations $j \geq 1$, the algorithm proceeds by setting

$$\theta^j = \theta^{j-1} + y_{i_j} \tilde{x}_{i_j}$$

Here $(\tilde{x}_{i_j}, y_{i_j}$ represents the input/output pair that is misclassified by θ^{j-1}. We hope that as we proceed with the iterations $j \geq 1$, the estimate θ^j converges to θ^* which separates the data. We want to find a finite j where this eventually happens which we can record as the upper bound for the algorithm's converge.

(a). Suppose that θ^* is normalized so that $\rho = min_i \mid \langle \theta^*, \tilde{x}_i \rangle \mid$ calculates the distance between the closest x_i in the training data to the hyperplane defined by θ^*. Please argue that

$$min_i \; y_i \langle \theta^*, \tilde{x}_i \rangle = \rho > 0$$

(b). Show by induction that

$$\langle \theta^j, \theta^* \rangle \geq \langle \theta^{j-1}, \theta^* \rangle + \rho$$

and conclude that $\langle \theta^j, \theta^* \rangle \geq j\rho$

(c). By using the fact that $\tilde{x}_{i_j}$ was misclassified by θ^{j-1} to show that

$$\|\theta^j\|^2 \leq \|\theta^{j-1}\|^2 + \|\tilde{x}_{i_j}\|^2$$

(d). Again, by induction, show that

$$\|\theta^j\|^2 \leq j(1 + R^2)$$

where $R = max_i \; \|x_i\|$, and $\|\|$ is just the Euclidean norm.

(e). Using Cauchy-Schwartz inequality, show (b) and (d) together implies that

$$j \leq \frac{(1 + R^2) \; \|\theta^*\|^2}{\rho^2}$$

Problem 2: [Implementation of PLA] Using the same synthetic data sets I provided in Homework 2, I would like you to implement LDA as described in our lecture note. If you are lazy, you can achieve this with a small modification of your code for stochastic gradient descent from Homework 2. Remember in our Lecture note, this resembles is also highlighted. As you have noticed with your previous homework, data set # 2 and #4 are not perfectly linearly separable in which case your PLA algorithm may run forever. It would be meaningful to put a stopping criteria for these data sets. Feel free to experiment with your stopping criteria, no harsh requirements on that.

Problem 3: [Growth Function] Please calculate the $m_{\mathcal{H}}(n)$ for the classifiers on $\mathbb{R}$ given below:

(a). $h(x) = sign(x - a)$ for some $a \in \mathbb{R}$ or $h(x) = -sign(x - a)$ for some $a \in \mathbb{R}$. This is the set of both positive and negative rays. You can find the positive ray example in Lecture 7 class notes.

(b). Also for the set of both positive and negative intervals described as below:

$$h(x) = \begin{cases} +1 & \text{for } \; x \in [a, b] \\ -1 & \text{otherwise} \end{cases}$$

or

$$h(x) = \begin{cases} -1 & \text{for } \; x \in [a, b] \\ +1 & \text{otherwise} \end{cases}$$

for some $a, b \in \mathbb{R}$.

Problem 4: [Break point] Consider classifiers in $\mathbb{R}^2$ described as, h such that for any $x \in \mathbb{R}^2$,

$$h(x) = \begin{cases} +1 & \text{if } \|x - c\| \leq r \\ -1 & \text{otherwise} \end{cases}$$

for some $c \in \mathbb{R}^2, \; r \in \mathbb{R}$.

(a). Show that this classifier shatters three-points data sets, i.e., $m_{\mathcal{H}}(3) = 8$.

(b). Show that $k = 4$ is a break point, i.e., $m_{\mathcal{H}}(4) < 16$.