EE 559 **Homework Week 6** Posted: Fri., 2/15/2019
Jenkins Due: Fri. 2/22/2019, 2:00 PM

1. For 2-class perceptron *with margin* algorithm, using basic sequential GD, fixed increment, prove convergence for linearly separable training data by modifying the perceptron convergence proof covered in class. You may write out the proof, or you may take the 3-page proof from lecture (included in this HW folder, as updated with corrections), and mark up the proof to show all changes as needed. If you mark up the existing proof, be sure to mark everything that needs changing (*e.g.*, if a change propagates through the proof, be sure to make all changes for a complete answer).

2. You are given the following training data points in three pattern classes S_1, S_2, and S_3:

$$\{(0,1,-1,2)\} \in S_1; \quad \{(1,1,1,1),(2,1,1,1)\} \in S_2; \quad \{(-1,1,0,-1)\} \in S_3$$

Note that in our notation (throughout this class), for convenience we can write (x_1,x_2,x_3,x_4) with commas, to denote a column vector (of dimension 4 in this case).

(a) Find linear discriminant functions that correctly classify the training data, using the multiclass Perceptron algorithm using maximal value method (given in Discussion Week 6). Use augmented space (so first augment the data). There are few enough iterations that this can be done by hand, or you may write code to do it if you prefer.

Use the following assumptions and starting point. Assume the data points have already been shuffled, so use the training data in the order given above. Use $\eta(i) = 1 \ \ \forall i$, and initial weight vectors:

$$\underline{w}^{(1)}(0) = -\underline{1}, \quad \underline{w}^{(2)}(0) = \underline{1}, \quad \underline{w}^{(3)}(0) = \underline{0}.$$

(b) From this 5-dimensional feature space, consider points that lie in the plane $\mathcal{P}$ defined by all $\underline{x}$ such that $\underline{x} = (1,x_1,x_2,0,0)$. Give the decision rule for points (x_1,x_2) that lie in this plane. Plot in 2-space, the decision boundaries and decision regions in plane $\mathcal{P}$.

3. Suppose you set up a training algorithm to use a modified MSE criterion:

$$J(\underline{w}) = \frac{1}{N}\left\|\underline{\underline{X}}\,\underline{w} - \underline{b}\right\|_2^2 + \lambda\left\|\underline{w}\right\|_2^2$$

in which the purpose of the new term is to prefer small $\left\|\underline{w}\right\|_2$ if $\lambda > 0$.

(a) Find $\nabla_{\underline{w}} J(\underline{w})$ using gradient relations.

(b) Find the optimal $\underline{w} = \hat{\underline{w}}$ by solving $\nabla_{\underline{w}} J(\underline{w}) = \underline{0}$. Compare your result to the pseudoinverse solution.

4. Starting from the MSE criterion function, derive a learning algorithm using the basic sequential gradient descent technique, as follows:

   (a) Find an expression for $J_n\left(\underline{w}\right)$ and from that derive an expression for $\nabla_{\underline{w}} J_n\left(\underline{w}\right)$.

   (b) Complete the derivation to get the sequential gradient descent algorithm based on MSE. Compare with the Widrow-Hoff learning algorithm given in lecture.