JHU Engineering for Professionals
Applied and Computational Mathematics
Data Mining 625.740

Homework for Module on Linear Discrimination

1. (a) Show that the distance from the hyperplane $g(\mathbf{x}) = \mathbf{w^T x} + w_0 = 0$ to the point $\mathbf{x}$ is $|g(\mathbf{x})|/||\mathbf{w}||$ by minimizing $||\mathbf{x} - \mathbf{x_q}||^2$ subject to the constraint $g(\mathbf{x_q}) = 0$.
   (b) Show that the projection of $\mathbf{x}$ onto the hyperplane is given by
$$\mathbf{x_p} = \mathbf{x} - \frac{g(\mathbf{x})}{||\mathbf{w}||^2}\mathbf{w}.$$

2. Let $\mathbf{x_1}, \ldots, \mathbf{x_n}$ be n q-dimensional samples and Q be any nonsingular positive definite $q \times q$ matrix. Show that the vector $\mathbf{x}$ that minimizes
$$\sum_{k=1}^{n} (\mathbf{x_k} - \mathbf{x})^T Q^{-1} (\mathbf{x_k} - \mathbf{x})$$
is the sample mean, $\bar{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x_k}$.

3. Consider a linear classifier with discriminant functions $g_i(\mathbf{x}) = \mathbf{w_i^T x} + w_{i0}, \quad i = 1, \ldots, c$. Show that the decision regions are convex by showing that if $\mathbf{x_1} \in \mathcal{R}_i$ and $\mathbf{x_2} \in \mathcal{R}_i$ then $\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \mathcal{R}_i$ if $0 \le \lambda \le 1$.

4. In the gradient descent algorithm, $\mathbf{a}_{k+1}$ is obtained from $\mathbf{a}_k$ by
$$\mathbf{a}_{k+1} = \mathbf{a}_k - \rho_k \nabla \mathbf{J}(\mathbf{a}_k),$$
where ρ_k is a positive scale factor that sets the step size. Consider the criterion function
$$J_q(\mathbf{a}) = \sum_{y \in \mathcal{Y}} (\mathbf{a}^T \mathbf{y} - b)^2$$
where $\mathcal{Y}(\mathbf{a})$ is the set of samples for which $\mathbf{a}^T \mathbf{y} \le b$. Suppose that $\mathbf{y}_1$ is the only sample in $\mathcal{Y}(\mathbf{a}_k)$. Show that $\nabla \mathbf{J}_q(\mathbf{a}_k) = 2(\mathbf{a}_k^T \mathbf{y}_1 - b)\mathbf{y}_1$ and that the matrix of second partial derivatives is given by $D = 2\mathbf{y}_1 \mathbf{y}_1^T$. Use this to show that when the optimal ρ_k is used in the gradient descent algorithm,
$$\mathbf{a}_{k+1} = \mathbf{a}_k + \frac{b - \mathbf{a}^T \mathbf{y}_1}{||\mathbf{y}_1||^2}\mathbf{y}_1.$$

5. Show that the partial derivatives of the functions $y_i = \exp(a_i)/\sum_j \exp(a_j)$ used in multiple class logistic discrimination are given by
$$\frac{\partial y_i}{\partial a_j} = y_i(\delta_{ij} - y_j)$$
where $\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & \text{otherwise.} \end{cases}$