EE E6887 (Statistical Pattern Recognition)
Solutions for homework 5

P.1 In a general quadratic discriminant function, we have

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d}\sum_{j=1}^{d} w_{ij} x_i x_j$$

Use the vector derivative rules we discussed in the class to find $\nabla_{\mathbf{x}}(g)$, namely the gradient of g with respect to $\mathbf{x}$, where $\mathbf{x} = [x_1, \ldots, x_d]^t$.

Answer:

$$\begin{aligned}\nabla_{\mathbf{x}}(g) &= \nabla_{\mathbf{x}}(\mathbf{w}^t\mathbf{x}) + \nabla_{\mathbf{x}}(\mathbf{x}^t\mathbf{W}\mathbf{x}) \\ &= \mathbf{w} + \mathbf{W}\mathbf{x} + \mathbf{W}^t\mathbf{x}\end{aligned}$$

P.2 The convex hull of a set of vectors $\mathbf{x}_i$, $i = 1, \ldots, n$ is the set of all vectors of the form

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i$$

where the coefficients α_i are nonnegative and sum to 1. One way to check whether two sets of vectors are linearly separable is to see if their convex hulls overlap.

Use the above method to check whether the following data sets are linearly separable.

$$\begin{aligned}\omega_1 : &\quad (1,2)^t \;\; (2,-4)^t \;\; (-3,-1)^t \\ \omega_2 : &\quad (2,4)^t \;\; (-1,-5)^t \;\; (5,0)^t\end{aligned}$$

Answer: Let $\mathbf{x}_1$ and $\mathbf{x}_2$ denote, respectively, the convex hulls of the samples from ω_1 and ω_2, and assume two parameters α_{11}, α_{12} for ω_1, and two parameters α_{21}, α_{22} for ω_2:

$$\begin{aligned}\mathbf{x}_1 &= (4\alpha_{11} + 5\alpha_{12} - 3,\;\; 3\alpha_{11} - 3\alpha_{12} - 1)^t \\ \mathbf{x}_2 &= (-3\alpha_{21} - 6\alpha_{22} + 5,\;\; 4\alpha_{21} - 5\alpha_{22})^t\end{aligned}$$

Assume that there is overlap between these two convex hulls. Let $\mathbf{x}$ be an arbitrary point in the overlap part. Then we have

$$\begin{cases} 4\alpha_{11} + 5\alpha_{12} - 3 = -3\alpha_{21} - 6\alpha_{22} + 5 \\ 3\alpha_{11} - 3\alpha_{12} - 1 = 4\alpha_{21} - 5\alpha_{22} \end{cases}$$

There are multiple solutions for this equation, and thus this overlap part exists. This means that the given two data sets are not linearly separable.

P.3 In the convergence proof of the Perceptron algorithm the scale factor α was taken to be $\frac{\beta^2}{2\gamma}$.

Using the notation of Section 5.5, show that if α is greater than $\frac{\beta^2}{2\gamma}$, the maximum number of correlation is given by:

$$k_0 = \frac{||\mathbf{a}_1 - \alpha\mathbf{a}||^2}{2\alpha\gamma - \beta^2}$$

Answer: From Theorem similar derivations as in Theorem 5.1, we can still get

$$||\mathbf{a}(k+1) - \alpha\hat{\mathbf{a}}||^2 \leq ||\mathbf{a}(k) - \alpha\hat{\mathbf{a}}||^2 - 2\gamma\alpha + \beta^2$$

When $\alpha > \frac{\beta^2}{2\gamma}$, $2\gamma\alpha - \beta^2 > \beta^2 > 0$, so the squared distance from $\mathbf{a}(k)$ to $\alpha\hat{\mathbf{a}}$ is still reduced by $2\gamma\alpha - \beta^2$ at each correction, and after k corrections we get

$$||\mathbf{a}(k+1) - \alpha\hat{\mathbf{a}}||^2 \leq ||\mathbf{a}(k) - \alpha\hat{\mathbf{a}}||^2 - k(2\gamma\alpha - \beta^2)$$

In total the sequence of corrections must terminate after no more than k_0 corrections, where k_0 is gotten from $||\mathbf{a}_1 - \alpha\hat{\mathbf{a}}||^2 - k(2\gamma\alpha - \beta^2) = 0$ as:

$$k_0 = \frac{||\mathbf{a}_1 - \alpha\hat{\mathbf{a}}||^2}{2\gamma\alpha - \beta^2}$$