Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Emilio Balda

Exercise 10

Friday, January 12, 2018

Problem 1. *(Discriminant Analysis)*

Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are n samples that have been clustered into g clusters $S_1, \ldots, S_g$. Let $\mathbf{X}_k$ be $\mathbf{X}_k = [\mathbf{x}_i]_{i \in S_k}$ and $\bar{\mathbf{x}}_k$ be the cluster centers for $k = 1, \ldots, g$, then prove the following statements.

a) For any S_l we have that

$$\sum_{i \in S_k} \left(\mathbf{x}_i \mathbf{x}_i^T - \overline{\mathbf{x}}_k \overline{\mathbf{x}}_k^T \right) = \sum_{i \in S_k} (\mathbf{x}_i - \overline{\mathbf{x}}_k)(\mathbf{x}_i - \overline{\mathbf{x}}_k)^T .$$

b) The matrix $\mathbf{W}$ corresponding to the sum of squares within groups (defined as $\mathbf{W} = \sum_{k=1}^{g} \mathbf{X}_k^T \mathbf{E}_k \mathbf{X}_k$) can be expressed as

$$\mathbf{W} = \sum_{k=1}^{g} \sum_{i \in S_k} (\mathbf{x}_i - \overline{\mathbf{x}}_k)(\mathbf{x}_i - \overline{\mathbf{x}}_k)^T .$$

Problem 2. *(Support Vector Machine with Only One Member per Class)* Let the dataset consist of only two points, $(\mathbf{x}_1, y_1 = +1)$ and $(\mathbf{x}_2, y_2 = -1)$. Find the SVM classifier and its parameters.

Problem 3. *(Support Vector Machine Margin)* Let the dataset consist of points, $(\mathbf{x}_i, y_i = +1)$, $i = 1, 2$ and $(\mathbf{x}_3, y_3 = -1)$. Suppose that these points are linearly separable.

a) Show that if these points are collinear, the maximum margin of the SVM classifier is obtained by the minimum of $\|\mathbf{x}_1 - \mathbf{x}_3\|$ and $\|\mathbf{x}_2 - \mathbf{x}_3\|$.

b) Discuss the case where the points are not collinear and argue why the margin cannot be as the above case.