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

Exercise 10
- Proposed Solution -

Friday, January 12, 2018

Solution of Problem 1

a)

$$\begin{aligned}\sum_{i\in S_k}(\mathbf{x}_i-\overline{\mathbf{x}}_k)(\mathbf{x}_i-\overline{\mathbf{x}}_k)^T &= \sum_{i\in S_k}\left(\mathbf{x}_i\mathbf{x}_i^T-2\mathbf{x}_i\overline{\mathbf{x}}_k^T+\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T\right)=\sum_{i\in S_k}\mathbf{x}_i\mathbf{x}_i^T-2\sum_{i\in S_k}\mathbf{x}_i\overline{\mathbf{x}}_k^T+\sum_{i\in S_k}\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T\\
&=\sum_{i\in S_k}\mathbf{x}_i\mathbf{x}_i^T-2\left(\sum_{i\in S_k}\mathbf{x}_i\right)\overline{\mathbf{x}}_k^T+|S_k|\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T\\
&=\sum_{i\in S_k}\mathbf{x}_i\mathbf{x}_i^T-2|S_k|\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T+|S_k|\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T\\
&=\sum_{i\in S_k}\mathbf{x}_i\mathbf{x}_i^T-|S_k|\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T=\sum_{i\in S_k}\mathbf{x}_i\mathbf{x}_i^T-\sum_{i\in S_k}\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T\\
&=\sum_{i\in S_k}\mathbf{x}_i\mathbf{x}_i^T-\overline{\mathbf{x}}_k\overline{\mathbf{x}}_k^T\end{aligned}$$

b) Without loss of generality let us consider the case where, for some k, we have that $S_k=\{1,\dots,m\}$, where $m=|S_k|$. Then, it follows

$$\mathbf{X}_k=[\mathbf{x}_1,\dots,\mathbf{x}_m]^T\,,\qquad\text{and}\qquad \mathbf{E}_k=\mathbf{I}_m-\frac{1}{m}\mathbf{1}_{m\times m}=\mathbf{I}_m-\frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T\,.$$

Therefore, we have that

$$\begin{aligned}\mathbf{X}_k^T\mathbf{E}_k\mathbf{X}_k&=\begin{bmatrix}\mathbf{x}_1&\dots&\mathbf{x}_m\end{bmatrix}\left(\mathbf{I}_m-\frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T\right)\begin{bmatrix}\mathbf{x}_1^T\\ \vdots\\ \mathbf{x}_m^T\end{bmatrix}\\
&=\begin{bmatrix}\mathbf{x}_1&\dots&\mathbf{x}_m\end{bmatrix}\begin{bmatrix}\mathbf{x}_1^T\\ \vdots\\ \mathbf{x}_m^T\end{bmatrix}-\frac{1}{m}\begin{bmatrix}\mathbf{x}_1&\dots&\mathbf{x}_m\end{bmatrix}\mathbf{1}_m\mathbf{1}_m^T\begin{bmatrix}\mathbf{x}_1^T\\ \vdots\\ \mathbf{x}_m^T\end{bmatrix}\\
&=\sum_{i=1}^m\mathbf{x}_i\mathbf{x}_i^T-\frac{1}{m}\left(\sum_{i=1}^m\mathbf{x}_i\right)\left(\sum_{i=1}^m\mathbf{x}_i\right)^T=\sum_{i=1}^m\mathbf{x}_i\mathbf{x}_i^T-\frac{1}{m}\left(m\bar{\mathbf{x}}_k\right)\left(m\bar{\mathbf{x}}_k\right)^T\\
&=\sum_{i=1}^m\left(\mathbf{x}_i\mathbf{x}_i^T-\bar{\mathbf{x}}_k\bar{\mathbf{x}}_k^T\right)=\sum_{i\in S_k}\left(\mathbf{x}_i\mathbf{x}_i^T-\bar{\mathbf{x}}_k\bar{\mathbf{x}}_k^T\right)\underset{\textbf{(a)}}{=}\sum_{i\in S_k}(\mathbf{x}_i-\overline{\mathbf{x}}_k)(\mathbf{x}_i-\overline{\mathbf{x}}_k)^T\,.\end{aligned}$$

Finally, this leads to

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

thus prooving the statement.

Solution of 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 $(\mathbf{x}_2, y_2 = -1)$. See that

$$\mathbf{a}^T\mathbf{x}_1 + b \geq 1$$
$$-\mathbf{a}^T\mathbf{x}_2 - b \geq 1.$$

Adding those inequalities provide

$$\mathbf{a}^T(\mathbf{x}_1 - \mathbf{x}_2) \geq 2 \implies \|\mathbf{a}\|_2\|\mathbf{x}_1 - \mathbf{x}_2\|_2 \geq 2.$$

Therefore $\|\mathbf{a}\|_2$, the objective function of the classifier achieves the minimum $\frac{2}{\|\mathbf{x}_1-\mathbf{x}_2\|_2}$ for

$$\mathbf{a} = \frac{2(\mathbf{x}_1 - \mathbf{x}_2)}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2}.$$

On the other side, we have:

$$\mathbf{a}^T\mathbf{x}_1 + b \geq 1 \implies \frac{2(\mathbf{x}_1^T\mathbf{x}_1 - \mathbf{x}_2^T\mathbf{x}_1)}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2} + b \geq 1$$
$$\implies b \geq \frac{(\mathbf{x}_2^T\mathbf{x}_2 - \mathbf{x}_1^T\mathbf{x}_1)}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2}.$$

and

$$-\mathbf{a}^T\mathbf{x}_2 - b \geq 1 \implies -\frac{2(\mathbf{x}_1^T\mathbf{x}_2 - \mathbf{x}_2^T\mathbf{x}_2)}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2} - b \geq 1$$
$$\implies b \leq \frac{(\mathbf{x}_2^T\mathbf{x}_2 - \mathbf{x}_1^T\mathbf{x}_1)}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2}.$$

which means that

$$b = \frac{\mathbf{x}_2^T\mathbf{x}_2 - \mathbf{x}_1^T\mathbf{x}_1}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2}.$$

The SVM classifier is given by:

$$\frac{2(\mathbf{x}_1^T\mathbf{x} - \mathbf{x}_2^T\mathbf{x})}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2} + \frac{\mathbf{x}_2^T\mathbf{x}_2 - \mathbf{x}_1^T\mathbf{x}_1}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2} \underset{y=-1}{\overset{y=1}{\gtrless}} 0.$$

However a bit of manipulation shows that:

$$\|\mathbf{x} - \mathbf{x}_2\|_2 \underset{y=-1}{\overset{y=1}{\gtrless}} \|\mathbf{x} - \mathbf{x}_1\|_2.$$

Solution of 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) First of all, we have:

$$\mathbf{a}^T\mathbf{x}_1 + b \geq 1$$
$$\mathbf{a}^T\mathbf{x}_2 + b \geq 1$$
$$-\mathbf{a}^T\mathbf{x}_3 - b \geq 1.$$

From these inequalities we obtain:

$$\mathbf{a}^T(\mathbf{x}_1 - \mathbf{x}_3) \geq 2 \implies \|\mathbf{a}\|_2\|\mathbf{x}_1 - \mathbf{x}_3\|_2 \geq 2.$$

$$\mathbf{a}^T(\mathbf{x}_2 - \mathbf{x}_3) \geq 2 \implies \|\mathbf{a}\|_2\|\mathbf{x}_2 - \mathbf{x}_3\|_2 \geq 2.$$

Therefore $\|\mathbf{a}\|_2$ should satisfy the all the previous inequalities and be strictly bigger that $\max(\frac{2}{\|\mathbf{x}_1-\mathbf{x}_3\|_2}, \frac{2}{\|\mathbf{x}_2-\mathbf{x}_3\|_2})$. Without loss of generality, assume this is obtained by $\mathbf{x}_1$. Consider the following choice:

$$\mathbf{a} = \frac{2(\mathbf{x}_1 - \mathbf{x}_3)}{\|\mathbf{x}_1 - \mathbf{x}_3\|_2^2}.$$

Although this achieves the minimum possible value of those inequalities, it might lead to a classifier that does not correctly classify the training points or violate the constraint above. From this $\mathbf{a}$, we can find the corresponding b as follows and then check to see if this choice can correctly classify the training data. We have:

$$\mathbf{a}^T\mathbf{x}_1 + b \geq 1 \implies \frac{2(\mathbf{x}_1^T\mathbf{x}_1 - \mathbf{x}_3^T\mathbf{x}_1)}{\|\mathbf{x}_1 - \mathbf{x}_3\|_2^2} + b \geq 1$$
$$\implies b \geq \frac{\mathbf{x}_3^T\mathbf{x}_3 - \mathbf{x}_1^T\mathbf{x}_1)}{\|\mathbf{x}_1 - \mathbf{x}_3\|_2^2}.$$

and

$$-\mathbf{a}^T\mathbf{x}_3 + b \geq 1 \implies -\frac{2(\mathbf{x}_1^T\mathbf{x}_3 - \mathbf{x}_3^T\mathbf{x}_3}{\|\mathbf{x}_1 - \mathbf{x}_3\|_2^2} - b \geq 1$$
$$\implies b \leq \frac{\mathbf{x}_3^T\mathbf{x}_3 - \mathbf{x}_1^T\mathbf{x}_1)}{\|\mathbf{x}_1 - \mathbf{x}_3\|_2^2}.$$

Therefore b is given by:

$$b = \frac{\mathbf{x}_3^T\mathbf{x}_3 - \mathbf{x}_1^T\mathbf{x}_1}{\|\mathbf{x}_1 - \mathbf{x}_3\|_2^2}.$$

Since b is obtained to satisfy two of the constraints, we need only to check the other one:

$$\mathbf{a}^T\mathbf{x}_2 + b \geq 1$$

This is equal to

$$\|\mathbf{x}_2 - \mathbf{x}_3\|_2 \geq \|\mathbf{x}_1 - \mathbf{x}_3\|_2 + \|\mathbf{x}_1 - \mathbf{x}_2\|_2.$$

But this is true for collinear points hence this is the correct choice and the margin is given by the distance of $\mathbf{x}_1$ and $\mathbf{x}_3$.

b) If the points are not collinear, the last inequality above can never be satisfied due to triangle inequality.