Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Exercise 11
- Proposed Solution -

Friday, January 18, 2019

Solution of Problem 1

(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

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

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:

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

and

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

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:

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

However a bit of manipulation shows that:

$$\|\mathbf{x} - \mathbf{x}_2\|_2 \gtrless_{y=-1}^{y=1} \|\mathbf{x} - \mathbf{x}_1\|_2.$$

Solution of Problem 2

(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:

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

From these inequalities we obtain:

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

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:

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

and

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

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

$$\|\mathbf{x}_2 - \mathbf{x}_3\|_2^2 \geq \|\mathbf{x}_1 - \mathbf{x}_3\|_2^2 + \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2.$$

But this is true for points that create obtuse triangle 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 form an acute triangle, the last inequality above can never be satisfied due to triangle inequality.

Solution of Problem 3

SVM equivalent formulations: We repeat the SVM problem as follows:

$$\arg \min_{\mathbf{a}\in\mathbb{R}^p, b\in\mathbb{R}} \frac{1}{2}\|\mathbf{a}\|^2 \quad \text{s.t.} \quad y_i(\mathbf{a}^T\mathbf{x}_i + b) \geq 1, \; i = 1, \ldots, n$$

We show first that this problem is equivalent to the following problems.

a) If the data is separable, there is a separating hyperplane correctly classifying the instances. For each of these hyperplanes $\mathbf{a}^T\mathbf{x} + b = 0$ with $\|\mathbf{a}\| = 1$, the distance of the training point $\mathbf{x}_i$ to the hyperplane is given by $|\mathbf{a}^T\mathbf{x}_i + b|$. The margin is given by $\min |\mathbf{a}^T\mathbf{x}_i + b|$. If the classification is correct, we have:

$$y_i(\mathbf{a}^T\mathbf{x}_i + b) = |\mathbf{a}^T\mathbf{x}_i + b|$$

Furthermore for any hyperplane with misclassification we have $\min y_i(\mathbf{a}^T\mathbf{x}_i + b) < 0$. Therefore any solution to the optimization problem

$$\arg \max_{\mathbf{a}\in\mathbb{R}^p, b\in\mathbb{R}, \|\mathbf{a}\|=1} \min_{i\in\{1,\ldots,n\}} y_i(\mathbf{a}^T\mathbf{x}_i + b)$$

will have zero classification error and maximum margin.

b) Note that the problem above looks for $(\mathbf{a}, b)$ such that the training error is zero and the margine is maximized. For a pair $(\mathbf{a}, b)$, the separating hyperplane is given by $\mathbf{a}^T\mathbf{x} + b = 0$. Assume $\|\mathbf{a}\| = 1$. The distance of the training point $\mathbf{x}_i$ to the hyperplane is given by $|\mathbf{a}^T\mathbf{x}_i + b|$. There for the smallest distance to the hyperplane is given by $\min_{i\in\{1,\ldots,n\}} |\mathbf{a}^T\mathbf{x}_i + b|$. Therefore this distance should be maximized while the training points are classified correctly, which means that:

$$\arg \max_{\mathbf{a}\in\mathbb{R}^p, b\in\mathbb{R}, \|\mathbf{a}\|=1} \min_{i\in\{1,\ldots,n\}} |\mathbf{a}^T\mathbf{x}_i + b| \quad \text{s.t.} \quad y_i(\mathbf{a}^T\mathbf{x}_i + b) > 0, \; i = 1, \ldots, n$$