Homework 4 Machine Learning

Oskar Hulthen 950801-1195 huoskar@student.chalmers.se
Alexander Branzell 931003-1977 alebra@student.chalmers.se

May 2018

1 Theoretical section

1.1 SVM

A hard margin state vector machine with the following training points:

$$+1: \ (2,\ 2),\ (4,\ 4),\ (0,\ 4)$$

$$-1: \ (0,\ 0),\ (2,\ 0),\ (0,\ 2)$$

Plotting these training points yield the following figure:

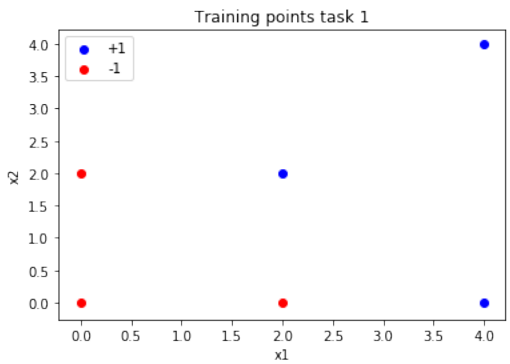

Figure 1: Plotted training points

From the figure we clearly see that to separate these two classes with a single line, that line has to travel in between the points (0, 2) and (2, 2) whilst also traveling in between the points (2, 0) and (4, 0). Assuming that the line travels exactly in between them we know that the points (1,2) and (3,0) are on the line, which gives us the equation:

$$x_2 = 3 - x_1 \rightarrow \mathbf{x_1 + x_2 - 3 = 0}$$

More generally following the equation for the margin for the four points we get:

$$(2,\ 2)\ (+1)\ :\ \frac{|w_1 \cdot 2 + w_2 \cdot 2 + b|}{\sqrt{w_1^2 + w_2^2}} = \gamma \tag{1}$$

$$(4,\ 0)\ (+1)\ :\ \frac{|w_1 \cdot 4 + b|}{\sqrt{w_1^2 + w_2^2}} = \gamma \tag{2}$$

$$(0,\ 2)\ (-1)\ :\ \frac{|w_2 \cdot 2 + b|}{\sqrt{w_1^2 + w_2^2}} = \gamma \tag{3}$$

$$(2,\ 0)\ (-1)\ :\ \frac{|w_1 \cdot 2 + b|}{\sqrt{w_1^2 + w_2^2}} = \gamma \tag{4}$$

From equation 3 = equation 4 since we also know that these two are on the same side of the decision boundary, we get that $w_1 = w_2 = w$. Thus implying that instead of the four above equations now get two, one for each of the classes (or sides of the decision boundary).

$$(+1) \ : \ \frac{|w \cdot 4 + b|}{\sqrt{2w^2}} = \gamma \tag{5}$$

$$(-1) \ : \ \frac{|w \cdot 2 + b|}{\sqrt{2w^2}} = \gamma \tag{6}$$

From equation 5 = equation 6 we get $|w \cdot 4 + b| \ = \ |w \cdot 2 + b|$, due to the fact that we know that these are on opposite sides of the decision boundary (one is negative and one is positive) we get the following:

$$\begin{aligned} &w \cdot 4 + b \ = \ -(w \cdot 2 + b) \\ &2b = -6w \\ &b = -3w \end{aligned} \tag{7}$$

From the now know relations between the variables ($w_1 = w_2$ and $b = -3w$, we can specify the scale by trying to classify some of the training points:

$$\begin{aligned} (2,\ 2) \ &: \ 2 \cdot w + 2 \cdot w - 3 \cdot w = 1 \\ (0,\ 4) \ &: \ 4 \cdot w - 3 \cdot w = 1 \\ (2,\ 0) \ &: \ 2 \cdot w - 3 \cdot w = -1 \\ (0,\ 2) \ &: \ 2 \cdot w - 3 \cdot w = -1 \end{aligned} \tag{8}$$

From the above equation system we get the solution that w = 1 and b = -3. In other words we get the following equation for the decision boundary hyperplane: $\mathbf{x_1 + x_2 - 3 = 0}$.

Furthermore the margin becomes:

$$2\gamma = 2 \cdot \frac{1}{\sqrt{w_1^2 + w_2^2}} = \frac{2}{\sqrt{2}} = \sqrt{2}$$

1.2 SVM cont'd

a)

For the primal formulation for this specific example we get:

$$\underset{\mathbf{w}}{argmin} \ \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$\begin{aligned} subject\ to \ : \ & 1 \cdot (\mathbf{w}^T(2,\ 2) + b) \geq 1 \\ & 1 \cdot (\mathbf{w}^T(4,\ 4) + b) \geq 1 \\ & 1 \cdot (\mathbf{w}^T(0,\ 4) + b) \geq 1 \\ & -1 \cdot (\mathbf{w}^T(0,\ 0) + b) \geq 1 \\ & -1 \cdot (\mathbf{w}^T(2,\ 0) + b) \geq 1 \\ & -1 \cdot (\mathbf{w}^T(0,\ 2) + b) \geq 1 \end{aligned}$$

b)

The optimal primal solution is calculated by substituting the conditions mentioned in section a). Adding the 5th and 6th constraint to the first constraint yields: $w_1 \geq 1$ and $w_2 \geq 1$. If the weights are then set to their lowest values (1) then it leads to $b \leq -3$ for every constraint to be satisfied. Due to us wanting to minimize the function (which doesnt contain b) we can conclude that the optimal solution is $w_1 = w_2 = 1$ and $b = -3$.

c)

The dual formulation for the svm with the lagrange multipliers α looks like this:

$$\underset{\alpha}{argmax} \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{N} \alpha_n \alpha_m t_n t_m \mathbf{x_n^T x_m}$$

$$subject\ to\ : \sum_{n=1}^{N} \alpha_n t_n = 0,\ \alpha_n \geq 0$$

Using the following values

$$x_1 = (2,\ 2),\ t_1 = 1$$

$$x_2 = (4,\ 4),\ t_2 = 1$$

$$x_3 = (0,\ 4),\ t_3 = 1$$

$$x_4 = (0,\ 0),\ t_4 = -1$$

$$x_5 = (2,\ 0),\ t_5 = -1$$

$$x_6 = (0,\ 2),\ t_6 = -1$$

Simplifying the above optimization function and constraint yielded:

$$max(\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5+\alpha_6-\frac{1}{2}(\alpha_1^2(8)+\alpha_1\alpha_2(16)-\alpha_1\alpha_5(8)-\alpha_1\alpha_6(8)+\alpha_2^2(32)+\alpha_2\alpha_3(32)$$

$$-\alpha_2\alpha_5(16) - \alpha_2\alpha_6(16) + \alpha_3^2(16) - \alpha_3\alpha_5(16) + \alpha_5^2(8) + \alpha_6^2(8)))$$

Subject to:

$$\alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 - \alpha_5 - \alpha_6 = 0$$

$$\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4,\ \alpha_5,\ \alpha_6 \geq 0$$

d)

Using the code present on the theoretical part of the notebook we reached the following solution:

$$\alpha \approx \begin{pmatrix} 1.5 \\ 0 \\ 1 \\ 0 \\ 1.75 \\ 0.75 \end{pmatrix}$$

We struggled to find a way to verify the solution, but we know that α_2 and α_4 will be zero in the solution, due to their constraints not being binding in the primal solution. By non binding means that their constraints in the primal solution are not tight, for example: $(1,\ 1)^T(4,\ 4) - 3 \geq 1$, where $(1,\ 1)^T(4,\ 4) - 3 = 5 \neq 1$. This also implies that the other data points are the support vectors for this specific problem: $(2,\ 2)$, $(0,\ 4)$, $(2,\ 0)$ and $(0,\ 2)$. Since this is the case for our solution, we believe it to be correct.