ECE421: Assignment #3

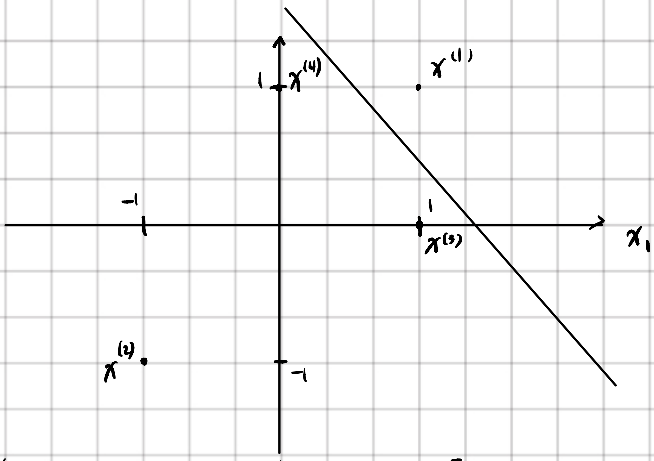
Won-Jin Kim
(1003611424)

Problem #1

Problem 1 The goal of this problem is to derive the parameters of a support vector machine with a hard margin. We will consider the following toy dataset of 4 points: $x^{(1)} = (1,1)^\top$, $x^{(2)} = (-1,-1)^\top$, $x^{(3)} = (1,0)^\top$, and $x^{(4)} = (0,1)^\top$. $x^{(1)}$ is the only point from the positive class, whereas $x^{(2)}$, $x^{(3)}$, and $x^{(4)}$ are from the negative class. Put another way, $y^{(1)} = 1$ and $y^{(2)} = y^{(3)} = y^{(4)} = -1$.

1. (2 points) Plot the data and draw the maximum-margin hyperplane. By inspecting the plot, give the equation of this hyperplane. You will use it to verify your answers at the end of this problem.

$\boxed{1}$



line passes through both $(0.5,1)^T$ and $(1,0.5)^T$

$$X_2 = \left(\frac{0.5-1}{1-0.5}\right)x_1 + 1.5$$

$$X_2 = -X_1 + 1.5$$

$$X_1 + X_2 - 1.5 = 0$$

2. (1 point) Which of the points are support vectors?

$\boxed{2}$ Support vectors of the maximum margin hyperplane are $x^{(1)} = (1,1)^T$, $x^{(3)} = (1,0)^T$, $x^{(4)} = (0,1)^T$

3. (1 point) Recall the optimization problem we defined for the SVM with a hard margin:

$$\min_{\vec{w},b} \frac{1}{2}\|\vec{w}\|^2 \ \texttt{s.t.} \quad \forall j(\vec{w}\cdot x^{(j)} + b)y^{(j)} \geq 1 \tag{1}$$

Write down the Lagrangian $L(\vec{w}, b, \vec{\alpha})$ corresponding to this problem, where $\vec{\alpha}$ is the vector of dual variables.

$\boxed{3}$

$$\mathcal{L}(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_j \alpha_j\left[(\vec{w}x^{(j)} + b)y^{(j)} - 1\right]$$

4. (2 points) We can therefore rewrite Equation 1 as follows:

$$\min_{\vec{w},b}\max_{\vec{\alpha}} L(\vec{w}, b, \vec{\alpha})$$

Because Slater's condition holds, this is equivalent to solving:

$$\max_{\vec{\alpha}}\min_{\vec{w},b} L(\vec{w}, b, \vec{\alpha}) \tag{2}$$

Given a fixed $\vec{\alpha}$, solve the inner (minimization) problem. You should obtain two relationships taking the form of

$$A = \sum_j B_j C_j \vec{D}_j$$

$$\sum_j E_j F_j = G$$

$\boxed{4}$ From previous problem.

$$\mathcal{L}(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_j \alpha_j\left[(\vec{w}x^{(j)} + b)y^{(j)} - 1\right]$$

$$\max_{\vec{\alpha}} \underbrace{\min_{\vec{w},b}}\ \mathcal{L}(\vec{w}, b, \vec{\alpha})$$

$$\frac{\partial \mathcal{L}(\vec{w}, b, \vec{\alpha})}{\partial w} = w + \sum_j \alpha_j x^{(j)} y^{(j)} = 0$$

$$\underline{w^* = \sum_j \alpha_j x^{(j)} y^{(j)}}$$

$$\frac{\partial \mathcal{L}(\vec{w}, b, \vec{\alpha})}{\partial b} = \underline{\sum_j \alpha_j y^{(j)} = 0}$$

fixing $\vec{\alpha}$, we get $\begin{cases} w^* = \sum_j \alpha_j x^{(j)} y^{(j)} \\ \sum_j \alpha_j y^{(j)} = 0 \end{cases}$

5. (2 points) Simplify Equation 2 with the result of question 4. Show that it takes the form of

$$\max_{\vec{\alpha}} \sum_j A_j - \frac{1}{2} \sum_j \sum_k B_j C_k D_j E_k \vec{F} \cdot \vec{G}$$

$\boxed{5}$

$\max_{\vec{\alpha}} \min_{\vec{w}, b} \mathcal{L}(w,b,\alpha)$

from previous questions we know that

$\mathcal{L}(w,b,\alpha) = \frac{1}{2}\|\vec{w}\|^2 - \sum_j \alpha_j [(\vec{w}x^{(j)}+b)y^{(j)} - 1]$

$w^* = \sum_j \alpha_j x^{(j)} y^{(j)}$ ——— (1)

$\sum_j \alpha_j y^{(j)} = 0$ ——— (2)

$\max_{\vec{\alpha}} \mathcal{L}(w,b,\alpha) = \max_{\vec{\alpha}} \Big(\frac{1}{2} \sum_j \sum_k y^{(j)} y^{(k)} x^{(j)} x^{(k)} \alpha_j \alpha_k - \sum_k \alpha_k \big((\sum_j \alpha_j x^{(j)} y^{(j)}) x^{(k)} + b \big) y^{(k)} + \sum_j \alpha_j \Big)$

$= \max_{\vec{\alpha}} \Big(\sum_j \alpha_j - \frac{1}{2} \sum_j \sum_k \alpha_j \alpha_k y^{(j)} y^{(k)} x^{(j)} x^{(k)} - b \cancel{\sum_j \alpha_j y^{(j)}} \Big)$

0 → according to equation (2)

$= \max_{\vec{\alpha}} \Big(\sum_j \alpha_j - \frac{1}{2} \sum_j \sum_k \alpha_j \alpha_k y^{(j)} y^{(k)} (x^{(j)T} x^{(k)}) \Big)$

6. (1 point) We now come back to our toy dataset. Because it has 4 trainings points, we have $j \in 1..4$, i.e., four dual variables. For points that are not support vectors, recall from class that the corresponding constraints in Equation 1 can be removed. Deduce the value of one of the dual variable α_j corresponding to one value of j (i.e., one constraint) which we will write j^* to not reveal the answer. Follow indexes introduced in the problem statement above when describing the dataset.

$\boxed{6}$ for $j=2$ $x^{(2)} = (-1, -1)^T$

it is not constrained from constraint where $j=2$.

$\therefore \alpha_2 = 0$

7. (1 point) Deduce from the second expression found in question 4 (i.e., $\sum_j E_j F_j = G$), a relationship between all remaining dual variables α_j for $j \neq j^*$

$\boxed{7}$ from Question 4

$\sum_j \alpha_j y^{(j)} = 0$

$\rightarrow \alpha_1 y^{(1)} + \alpha_2 y^{(2)} + \alpha_3 y^{(3)} + \alpha_4 y^{(4)}$

$= \alpha_1 - \alpha_2 - \alpha_3 - \alpha_4 = 0$

$\downarrow \alpha_2 = 0$

$= \alpha_1 - \alpha_3 - \alpha_4 = 0$

8. (3 points) Solving the optimization problem from question 5, and using question 7 in addition, find values for all dual variables α_j for $j \in 1..4$.

$\boxed{8}$

$$\alpha_1, \alpha_3, \alpha_3 = \max_{\alpha_1,\alpha_3,\alpha_4} \left(\sum_{j=1}^{4} \alpha_j - \frac{1}{2} \sum_{j=1}^{4} \sum_{k=1}^{4} \alpha_j \alpha_k y^{(j)} y^{(k)} (x^{(j)T} x^{(k)}) \right)$$

Compute pairwise dot product between training example.

$x^{(1)T}x^{(1)} = 2$

$x^{(1)T}x^{(3)} = 1$

$x^{(1)T}x^{(4)} = 1$

$x^{(3)T}x^{(3)} = 1$

$x^{(3)T}x^{(4)} = 0$

$x^{(4)T}x^{(4)} = 1$

$x^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x^{(3)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad x^{(4)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$y^{(1)} = 1 \quad y^{(3)} = y^{(4)} = -1$

Expand $\mathcal{L}(w^*, b^*, \alpha)$

$$\Rightarrow \max_{\alpha_1,\alpha_3,\alpha_4} -\frac{1}{2}\left(2\alpha_1^2 - \alpha_1\alpha_3 - \alpha_1\alpha_4 + \alpha_3^2 + \alpha_4^2\right) + \alpha_1 + \alpha_3 + \alpha_4$$

we now have dual optimal problem.

$$\begin{cases} \max_{\alpha_1 \alpha_3 \alpha_4} -\frac{1}{2}(2\alpha_1^2 - \alpha_1\alpha_3 - \alpha_1\alpha_4 + \alpha_3^2 + \alpha_4^2) + \alpha_1 + \alpha_3 + \alpha_4 \\ \alpha_1 - \alpha_3 - \alpha_4 = 0 \leftarrow \text{from previous question} \end{cases}$$

$\downarrow$ eliminate one variable

$\alpha_4 = \alpha_1 - \alpha_3$

$$\max_{\alpha_1,\alpha_3} -\frac{1}{2}\left(2\alpha_1^2 - \alpha_1\alpha_3 - \alpha_1(\alpha_1 - \alpha_3) + \alpha_3^2 + (\alpha_1 - \alpha_3)^2\right) + \alpha_1 + \alpha_3 + (\alpha_1 - \alpha_3)$$

$$= \max_{\alpha_1,\alpha_3} -\frac{1}{2}\left(2\alpha_1^2 - \alpha_1^2 - \alpha_1\alpha_3 + \alpha_1\alpha_3 + \alpha_3^2 + \alpha_1^2 - 2\alpha_1\alpha_3 + \alpha_3^2\right) + 2\alpha_1$$

$$= \max_{\alpha_1,\alpha_3} -\frac{1}{2}\left(2\alpha_1^2 - 2\alpha_1\alpha_3 + 2\alpha_3^2\right) + 2\alpha_1$$

Compute partial derivative wrt α & set them to 0.

$$\frac{\partial \mathcal{L}}{\partial \alpha_1} = -2\alpha_1 + \alpha_3 + 2 = 0 \quad \text{———} \quad (*)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_3} = \alpha_1 - 2\alpha_3 = 0 \rightarrow \alpha_1 = 2\alpha_3 \quad \text{———} \quad (**)$$

Substitute $(**)$ into $(*)$

$$-2(2\alpha_3) + \alpha_3 + 2 = 0$$

$$-3\alpha_3 = -2$$

$$\alpha_3 = \frac{2}{3}$$

Substitute this result to $(**)$

$$\alpha_1 = 2\left(\frac{2}{3}\right) = \frac{4}{3}$$

$$\alpha_4 = \alpha_1 - \alpha_3 = \frac{2}{3} = \alpha_3$$

9. (1 point) Deduce the numerical values of the two components of $\vec{w}$.

$\boxed{9}$

$$w^* = \sum_{i=1}^{4} \alpha_i y_i x^{(i)}$$

$$w^* = \frac{4}{3}(1)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{2}{3}(-1)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{2}{3}(-1)\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{4}{3} - \frac{2}{3} \\ \frac{4}{3} - \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

10. (2 points) Through an analysis of constraints in Equation 1 which are tight, deduce the numerical value of b. Hint: recall the role of support vectors here.

$\boxed{10}$ bias can be computed from any support vectors

$$y^{(1)}(w^{(1)T}x^{(1)} + b) = 1$$

$$b = 1 - w^{(1)T}x^{(1)} = 1 - \begin{bmatrix} \frac{2}{3} & \frac{2}{3} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 - \frac{4}{3} = -\frac{1}{3}$$