ECE421 Problem Set 3

Micol Altomare

October 11, 2023

1 SVM with Hard Margin

1.1

By inspection, the equation of the maximum-margin hyperplane is $\boxed{x_2 = -x_1 + 1.5}$.

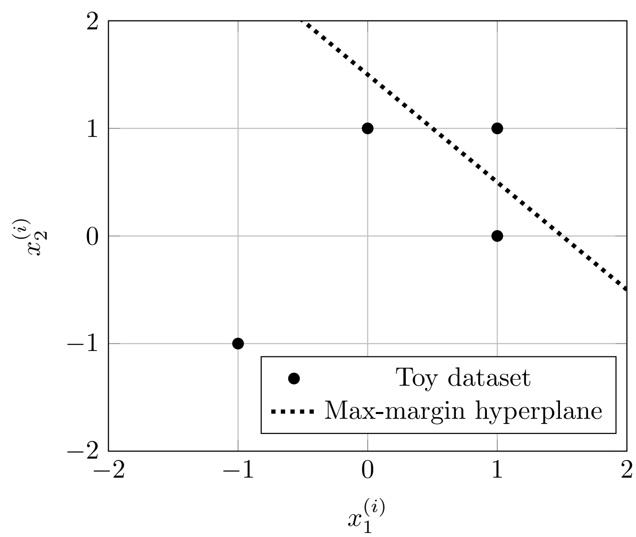

Figure 1: Scatter plot of the toy dataset.

1.2

$\boxed{x^{(1)}, x^{(3)}, x^{(4)}}$ are support vectors since they are closest to the hyperplane, and since their margin constraint will be 1 (ie. they are on the margin).

1.3

Recall the Lagrange function:

$$L(\bar{x}, \lambda) = f_0(\bar{x}) + \sum_{i=1} \lambda_i f_i(\bar{x}), \lambda_i \geq 0$$

Given optimization problem:

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

$$\begin{aligned}
\text{Objective function: } f_0(\bar{x}) &= \frac{1}{2}||\vec{w}||^2 \\
\text{Constraint: } &(\vec{w}\cdot\vec{x}^{(j)}+b)y^{(j)} \geq 1\forall j \\
&\implies (\vec{w}\cdot\vec{x}^{(j)}+b)y^{(j)} - 1 \geq 0\forall j \\
&\implies \boxed{L(\vec{w},b,\vec{\alpha}) = \frac{1}{2}||\vec{w}||^2 - \sum_j \alpha_j[(\vec{w}\cdot\vec{x}^{(j)}+b)y^{(j)}-1]}
\end{aligned}$$

1.4

$$\min_{\vec{w},b} L(\vec{w},b,\vec{\alpha}) \implies \frac{\partial}{\partial\vec{w}}L(\vec{w},b,\vec{\alpha}) = 0, \frac{\partial}{\partial b}L(\vec{w},b,\vec{\alpha}) = 0$$

Given the result from **1.3**,

$$\begin{aligned}
\frac{\partial}{\partial\vec{w}}L(\vec{w},b,\vec{\alpha}) &= \vec{w} - \sum_j \alpha_j y^{(j)}\vec{x}^{(j)} = 0 \\
&\implies \boxed{\vec{w} = \sum_j \alpha_j y^{(j)}\vec{x}^{(j)}} \\
\frac{\partial}{\partial b}L(\vec{w},b,\vec{\alpha}) &= -\sum_j \alpha_j y^{(j)} = 0 \\
&\implies \boxed{0 = \sum_j \alpha_j y^{(j)}}
\end{aligned}$$

1.5

$$\begin{aligned}
&\max_{\vec{\alpha}} \min_{\vec{w},b} L(\vec{w},b,\vec{\alpha}) \\
&= \max_{\vec{\alpha}} \min_{\vec{w},b} \left(\frac{1}{2}||\vec{w}||^2 - \sum_j \alpha_j[(\vec{w}\cdot\vec{x}^{(j)}+b)y^{(j)}-1]\right) \\
&= \max_{\vec{\alpha}} \left(\frac{1}{2}||\vec{w}||^2 - \sum_j (\alpha_j\vec{w}\cdot\vec{x}^{(j)}y^{(j)} + \alpha_j b y^{(j)} - \alpha_j)\right)
\end{aligned}$$

Given the results from **1.4**,

$$\begin{aligned}
&= \max_{\vec{\alpha}} \left(\frac{1}{2}||\vec{w}||^2 - \sum_j \alpha_j\vec{x}^{(j)}y^{(j)}\sum_k \alpha_k y^{(k)}\vec{x}^{(k)} + \sum_j \alpha_j\right) \\
&= \max_{\vec{\alpha}} \left(\frac{1}{2}\sum_j\sum_k \alpha_j\alpha_k y^{(j)}y^{(k)}(\vec{x}^{(j)}\cdot\vec{x}^{(k)}) - \sum_j\sum_k \alpha_j\alpha_k y^{(j)}y^{(k)}(\vec{x}^{(j)}\cdot\vec{x}^{(k)}) + \sum_j \alpha_j\right) \\
&= \boxed{\max_{\vec{\alpha}} \left(\sum_j \alpha_j - \frac{1}{2}\sum_j\sum_k \alpha_j\alpha_k y^{(j)}y^{(k)}(\vec{x}^{(j)}\cdot\vec{x}^{(k)})\right)}
\end{aligned}$$

1.6

Original constraint:

$$(\vec{w} \cdot \vec{x}^{(j)} + b)y^{(j)} \geq 1 \; \forall j$$

From **1.1**, $\vec{x}^{(2)}$ is the only non-support vector. From the Lagrange function, if the original constraint does not apply for non-support vectors, then $\sum_j \alpha_j(\vec{w}\cdot\vec{x}^{(j)}+b)y^{(j)}-1$ must be zero. Thus, α_j corresponding to the non-support vector must be zero. $\boxed{\therefore \alpha_{j*} = 0 \text{ when } j^* = 2}$.

1.7

Using the second result from **1.4** and the dataset,

$$\begin{aligned}
0 &= \sum_{j=1}^{4} \alpha_j y^{(j)} \\
&= \alpha_1 y^{(1)} + \alpha_2 y^{(2)} + \alpha_3 y^{(3)} + \alpha_4 y^{(4)} \\
&= \alpha_1(1) + 0(-1) + \alpha_3(-1) + \alpha_4(-1) \\
&= \alpha_1 - \alpha 3 - \alpha_4 \\
\implies & \boxed{\alpha_1 = \alpha_3 + \alpha_4}
\end{aligned}$$

1.8

To solve the optimization problem from **1.5**, we can expand the optimization problem and utilize the dataset points, positive/negative class assignments and the deduced dual value relationships: $\alpha_2 = 0, \alpha_1 = \alpha_3 + \alpha_4 \implies \alpha_3 = \alpha_1 - \alpha_4, \alpha_4 = \alpha_1 - \alpha_3$.

$$\begin{aligned}
&\max_{\vec{\alpha}} \left(\sum_{j=1}^{4} \alpha_j - \frac{1}{2} \sum_{j=1}^{4} \sum_{k=1}^{4} \alpha_j \alpha_k y^{(j)} y^{(k)} (\vec{x}^{(j)} \cdot \vec{x}^{(k)}) \right) \\
&= \max_{\vec{\alpha}} \left(\alpha_1 + \alpha_3 + \alpha_4 - \frac{1}{2} \sum_{j=1}^{4} \alpha_j y^{(j)} \sum_{k=1}^{4} \alpha_k y^{(k)} (\vec{x}^{(j)} \cdot \vec{x}^{(k)}) \right) \\
&= \max_{\vec{\alpha}} \left(\alpha_1 + \alpha_3 + \alpha_4 - \frac{1}{2} (\alpha_1(2\alpha_1 - \alpha_3 - \alpha_4) + 0 - \alpha_3(\alpha_1 - \alpha_3) - \alpha_4(\alpha_1 - \alpha_4)) \right) \\
&= \max_{\vec{\alpha}} \left(2\alpha_3 + 2\alpha_4 - \frac{1}{2} \left(2\alpha_1^2 - \alpha_1\alpha_3 - \alpha_1\alpha_4 - \alpha_3\alpha_1 + \alpha_3^2 - \alpha_4\alpha_1 + \alpha_4^2\right) \right) \\
&= \max_{\vec{\alpha}} \left(2\alpha_3 + 2\alpha_4 - \frac{1}{2} \left(2\alpha_1^2 - 2\alpha_1\alpha_3 - 2\alpha_1\alpha_4 + \alpha_3^2 + \alpha_4^2\right) \right) \\
&= \max_{\vec{\alpha}} \left(2\alpha_3 + 2\alpha_4 - \frac{1}{2} \left(\alpha_3^2 + \alpha_4^2\right) \right)
\end{aligned}$$

Note: $2\alpha_1^2 - 2\alpha_1\alpha_3 - 2\alpha_1\alpha_4 = 2\alpha_1^2 - 2\alpha_1(\alpha_1 - \alpha_4) - 2\alpha_1(\alpha_1 - \alpha_3) = 2\alpha_1^2 - 2\alpha_1^2 + 2\alpha_1(\alpha_3 + \alpha_4) - 2\alpha_1^2 = 0$

Now, we may take the partial derivatives of the simplified optimization problem and set them to zero:

$$\begin{aligned}
\frac{\partial}{\partial \alpha_3}\left(2\alpha_3 + 2\alpha_4 - \frac{1}{2}\left(\alpha_3^2 + \alpha_4^2\right)\right) &= 2 - \alpha_3 = 0 \\
&\implies \alpha_3 = 2 \\
\frac{\partial}{\partial \alpha_4}\left(2\alpha_3 + 2\alpha_4 - \frac{1}{2}\left(\alpha_3^2 + \alpha_4^2\right)\right) &= 2 - \alpha_4 = 0 \\
&\implies \alpha_4 = 2 \\
&\quad\;\; \alpha_1 = \alpha_3 + \alpha_4 \\
&\implies \alpha_1 = 4
\end{aligned}$$

$$\boxed{\therefore \vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \\ 2 \end{bmatrix}}$$

1.9

Using the first result from **1.4**, **1.8**, and the dataset,

$$\begin{aligned}
\vec{w} &= \sum_j \alpha_j y^{(j)} \vec{x}^{(j)} \\
&= \alpha_1 y^{(1)} \vec{x}^{(1)} + \alpha_2 y^{(2)} \vec{x}^{(2)} + \alpha_3 y^{(3)} \vec{x}^{(3)} + \alpha_4 y^{(4)} \vec{x}^{(4)} \\
&= 4 \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 4 - 2 - 0 \\ 4 - 0 - 2 \end{bmatrix}
\end{aligned}$$

$$\boxed{\vec{w} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}}$$

1.10

Original constraint:

$$(\vec{w} \cdot \vec{x}^{(j)} + b) y^{(j)} \geq 1 \; \forall j$$

For support vectors, the constraint is tight:

$$(\vec{w} \cdot \vec{x}^{(j)} + b) y^{(j)} = 1 \; \forall j$$

Any of the support vectors can be used to determine b, like $x^{(1)}$:

$$\begin{aligned}
(\vec{w} \cdot \vec{x}^{(1)} + b) y^{(1)} &= 1 \\
\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b\right) 1 &= 1 \\
4 + b &= 1 \\
\implies \boxed{b = -3}
\end{aligned}$$