Exercise of Supervised Learning: SVM Part 1

Yawei Li

yawei.li@stat.uni-muenchen.de

December 8, 2023

Exercise 1: Soft Margin Classifier

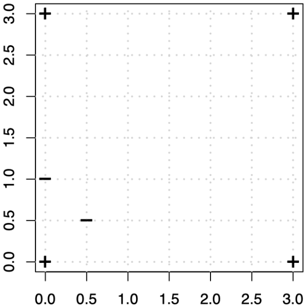
The primal optimization problem for the two-class soft margin SVM classification is given by

$$\min_{\boldsymbol{\theta},\theta_0,\zeta^{(i)}} \frac{1}{2}||\boldsymbol{\theta}||^2 + \sum_{i=1}^{n} \zeta^{(i)}$$

$$\text{s.t.:} \quad y^{(i)}(\boldsymbol{\theta}^T\mathbf{x}^{(i)} + \theta_0) \geq 1 - \zeta^{(i)},$$

$$\zeta^{(i)} \geq 0, \qquad \forall i = 1, \ldots, n.$$

(a) Add the decision boundary to the figure for $\hat{\boldsymbol{\theta}} = (1,1)^T, \hat{\theta}_0 = -2$. (NB: This is the approximate optimum for $C = 10$).



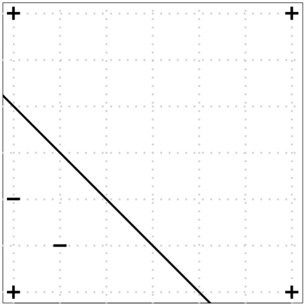
Solution to Exercise 1 (a)

The hyperplane is given by:

$$\theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \theta_0 = 0$$

Plugging in the values for the θs and solving for x_2, we get the decision boundary:

$$x_2 = -x_1 + 2$$



Draw this figure on whiteboard.

Exercise 1 (b)

(b) Identify the coordinates of the support vector(s) and compute the values of their slack variable $\zeta^{(i)}$.

Solution to Exercise 1 (b)

To determine which points are support vectors, we will use the constraint:

$$y^{(i)}(\mathbf{x}^{(i)}\hat{\boldsymbol{\theta}} + \hat{\theta}_0) \geq 1 - \zeta^{(i)}$$

- $(0,0)$: $1(0+0-2) = -2 \geq 1 - \zeta^{(1)} \rightarrow \zeta^{(1)} \geq 3$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 3$.
- $(0.5,0.5)$: $-1(0.5+0.5-2) = 1 \geq 1 - \zeta^{(2)} \rightarrow \zeta^{(2)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,1)$: $-1(0+1-2) = 1 \geq 1 - \zeta^{(3)} \rightarrow \zeta^{(3)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,3)$: $1(0+3-2) = 1 \geq 1 - \zeta^{(4)} \rightarrow \zeta^{(4)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,0)$: $1(3+0-2) = 1 \geq 1 - \zeta^{(5)} \rightarrow \zeta^{(5)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,3)$: $1(3+3-2) = 4 \geq 1 - \zeta^{(6)} \rightarrow \zeta^{(6)} \geq -3$, $\rightsquigarrow$ **Not** a support vector.

Solution to Exercise 1 (b)

To determine which points are support vectors, we will use the constraint:

$$y^{(i)}(\mathbf{x}^{(i)}\hat{\boldsymbol{\theta}} + \hat{\theta}_0) \geq 1 - \zeta^{(i)}$$

- $(0,0)$: $1(0+0-2) = -2 \geq 1 - \zeta^{(1)} \rightarrow \zeta^{(1)} \geq 3$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 3$.
- $(0.5, 0.5)$: $-1(0.5+0.5-2) = 1 \geq 1 - \zeta^{(2)} \rightarrow \zeta^{(2)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,1)$: $-1(0+1-2) = 1 \geq 1 - \zeta^{(3)} \rightarrow \zeta^{(3)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,3)$: $1(0+3-2) = 1 \geq 1 - \zeta^{(4)} \rightarrow \zeta^{(4)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,0)$: $1(3+0-2) = 1 \geq 1 - \zeta^{(5)} \rightarrow \zeta^{(5)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,3)$: $1(3+3-2) = 4 \geq 1 - \zeta^{(6)} \rightarrow \zeta^{(6)} \geq -3$, $\rightsquigarrow$ **Not** a support vector.

Solution to Exercise 1 (b)

To determine which points are support vectors, we will use the constraint:

$$y^{(i)}(\mathbf{x}^{(i)}\hat{\boldsymbol{\theta}} + \hat{\theta}_0) \geq 1 - \zeta^{(i)}$$

- $(0, 0)$: $1(0 + 0 - 2) = -2 \geq 1 - \zeta^{(1)} \rightarrow \zeta^{(1)} \geq 3$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 3$.
- $(0.5, 0.5)$: $-1(0.5 + 0.5 - 2) = 1 \geq 1 - \zeta^{(2)} \rightarrow \zeta^{(2)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0, 1)$: $-1(0 + 1 - 2) = 1 \geq 1 - \zeta^{(3)} \rightarrow \zeta^{(3)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0, 3)$: $1(0 + 3 - 2) = 1 \geq 1 - \zeta^{(4)} \rightarrow \zeta^{(4)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3, 0)$: $1(3 + 0 - 2) = 1 \geq 1 - \zeta^{(5)} \rightarrow \zeta^{(5)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3, 3)$: $1(3 + 3 - 2) = 4 \geq 1 - \zeta^{(6)} \rightarrow \zeta^{(6)} \geq -3$, $\rightsquigarrow$ **Not** a support vector.

Solution to Exercise 1 (b)

To determine which points are support vectors, we will use the constraint:

$$y^{(i)}(\mathbf{x}^{(i)}\hat{\boldsymbol{\theta}} + \hat{\theta}_0) \geq 1 - \zeta^{(i)}$$

- $(0,0)$: $1(0+0-2) = -2 \geq 1 - \zeta^{(1)} \rightarrow \zeta^{(1)} \geq 3$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 3$.
- $(0.5,0.5)$: $-1(0.5+0.5-2) = 1 \geq 1 - \zeta^{(2)} \rightarrow \zeta^{(2)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,1)$: $-1(0+1-2) = 1 \geq 1 - \zeta^{(3)} \rightarrow \zeta^{(3)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,3)$: $1(0+3-2) = 1 \geq 1 - \zeta^{(4)} \rightarrow \zeta^{(4)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,0)$: $1(3+0-2) = 1 \geq 1 - \zeta^{(5)} \rightarrow \zeta^{(5)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,3)$: $1(3+3-2) = 4 \geq 1 - \zeta^{(6)} \rightarrow \zeta^{(6)} \geq -3$, $\rightsquigarrow$ **Not** a support vector.

Solution to Exercise 1 (b)

To determine which points are support vectors, we will use the constraint:

$$y^{(i)}(\mathbf{x}^{(i)}\hat{\boldsymbol{\theta}} + \hat{\theta}_0) \geq 1 - \zeta^{(i)}$$

- $(0,0)$: $1(0+0-2) = -2 \geq 1 - \zeta^{(1)} \rightarrow \zeta^{(1)} \geq 3$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 3$.
- $(0.5,0.5)$: $-1(0.5+0.5-2) = 1 \geq 1 - \zeta^{(2)} \rightarrow \zeta^{(2)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,1)$: $-1(0+1-2) = 1 \geq 1 - \zeta^{(3)} \rightarrow \zeta^{(3)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,3)$: $1(0+3-2) = 1 \geq 1 - \zeta^{(4)} \rightarrow \zeta^{(4)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,0)$: $1(3+0-2) = 1 \geq 1 - \zeta^{(5)} \rightarrow \zeta^{(5)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,3)$: $1(3+3-2) = 4 \geq 1 - \zeta^{(6)} \rightarrow \zeta^{(6)} \geq -3$, $\rightsquigarrow$ **Not** a support vector.

Solution to Exercise 1 (b)

To determine which points are support vectors, we will use the constraint:

$$y^{(i)}(\mathbf{x}^{(i)}\hat{\boldsymbol{\theta}} + \hat{\theta}_0) \geq 1 - \zeta^{(i)}$$

- $(0,0)$: $1(0+0-2) = -2 \geq 1 - \zeta^{(1)} \rightarrow \zeta^{(1)} \geq 3$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 3$.
- $(0.5, 0.5)$: $-1(0.5+0.5-2) = 1 \geq 1 - \zeta^{(2)} \rightarrow \zeta^{(2)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,1)$: $-1(0+1-2) = 1 \geq 1 - \zeta^{(3)} \rightarrow \zeta^{(3)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,3)$: $1(0+3-2) = 1 \geq 1 - \zeta^{(4)} \rightarrow \zeta^{(4)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,0)$: $1(3+0-2) = 1 \geq 1 - \zeta^{(5)} \rightarrow \zeta^{(5)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,3)$: $1(3+3-2) = 4 \geq 1 - \zeta^{(6)} \rightarrow \zeta^{(6)} \geq -3$, $\rightsquigarrow$ **Not** a support vector.

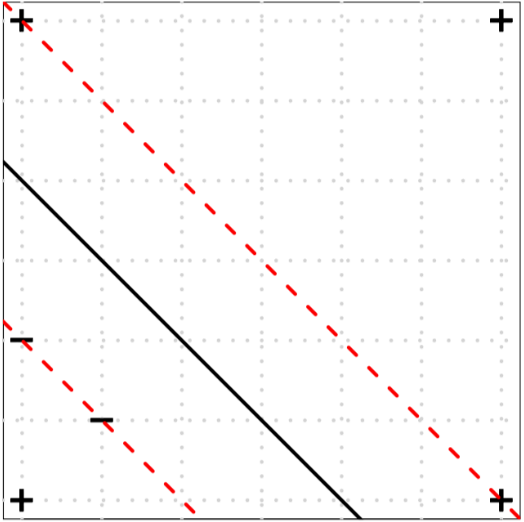
Solution to Exercise 1 (b)

To determine which points are support vectors, we will use the constraint:

$$y^{(i)}(\mathbf{x}^{(i)}\hat{\boldsymbol{\theta}} + \hat{\theta}_0) \geq 1 - \zeta^{(i)}$$

- $(0,0)$: $1(0+0-2) = -2 \geq 1 - \zeta^{(1)} \rightarrow \zeta^{(1)} \geq 3$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 3$.
- $(0.5, 0.5)$: $-1(0.5+0.5-2) = 1 \geq 1 - \zeta^{(2)} \rightarrow \zeta^{(2)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,1)$: $-1(0+1-2) = 1 \geq 1 - \zeta^{(3)} \rightarrow \zeta^{(3)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(0,3)$: $1(0+3-2) = 1 \geq 1 - \zeta^{(4)} \rightarrow \zeta^{(4)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,0)$: $1(3+0-2) = 1 \geq 1 - \zeta^{(5)} \rightarrow \zeta^{(5)} \geq 0$, $\rightsquigarrow$ Support vector with slack variable $\zeta^{(i)} = 0$.
- $(3,3)$: $1(3+3-2) = 4 \geq 1 - \zeta^{(6)} \rightarrow \zeta^{(6)} \geq -3$, $\rightsquigarrow$ **Not** a support vector.

Solution to Exercise 1 (b): Continued



Exercise 1 (c)

(c) Compute the Euclidean distance of the non-margin-violating support vector(s) to the decision boundary.

Solution to Exercise 1 (c)

We can use $\mathbf{x}^{(i)} = (0.5, 0.5)^T$:

$$d(f, \mathbf{x}^{(i)}) = \frac{y^{(i)} f(\mathbf{x}^{(i)})}{||\boldsymbol{\theta}||_2} = \frac{-1(0.5 + 0.5 - 2)}{\sqrt{2}} = \frac{1}{\sqrt{2}}$$

The distance is the same for all non-margin-violating support vectors.

Exercise 1 (d)

(d) What needs to be changed in the plot such that a hard margin SVM results into the same decision boundary?

Solution to Exercise 1 (d)

Some alternatives are:

- Convert the $(0,0)^T$ into a negative class.
- Move the $(0,0)^T$ to $(2,2)^T$.
- Delete $(0,0)^T$.

Solution to Exercise 1 (d)

Some alternatives are:

- Convert the $(0,0)^T$ into a negative class.
- Move the $(0,0)^T$ to $(2,2)^T$.
- Delete $(0,0)^T$.

Solution to Exercise 1 (d)

Some alternatives are:

- Convert the $(0,0)^T$ into a negative class.
- Move the $(0,0)^T$ to $(2,2)^T$.
- Delete $(0,0)^T$.

Solution to Exercise 1 (d)

Some alternatives are:

- Convert the $(0, 0)^T$ into a negative class.
- Move the $(0, 0)^T$ to $(2, 2)^T$.
- Delete $(0, 0)^T$.

Exercise 2: Optimization

Write your own stochastic subgradient descent routine to solve the soft-margin SVM in the primal formulation.

Hints:

- Use the regularized-empirical-risk-minimization formulation, i.e., an optimization criterion without constraints.
- No kernels, just a linear SVM.
- Compare your implementation with an existing implementation (e.g. `kernallab` in `R`. Are your results similar? Note that you might have to switch off the automatic data scaling in the already existing implementation.

Solution: show the standard solution.