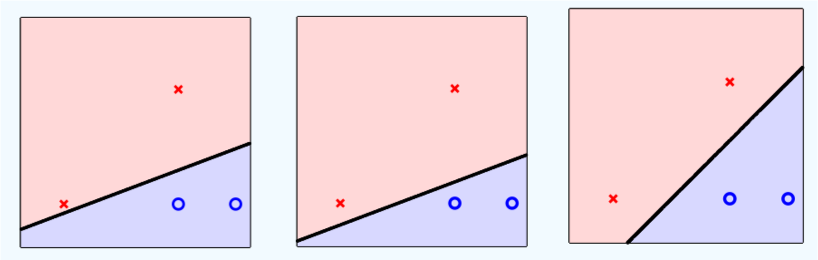
Agenda – SVM

- Fomulação para problemas de classificação binária
- Fronteiras lineares
- SVM hard-margin é QP (caso linearmente separável)
- Dual do SVM hard-margin reduz-se a um QP
- SVM soft-margin (caso não necessariamente linearmente separável)
- Kernel trick (não linearização)

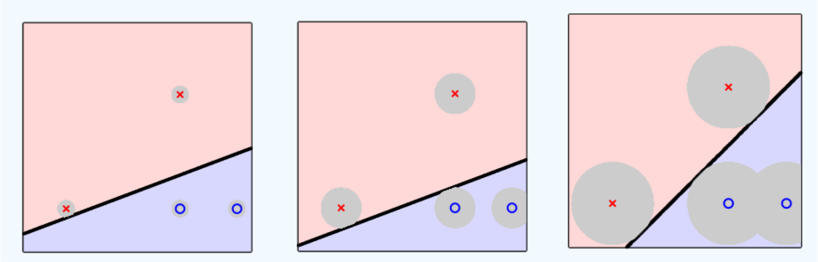
Linear decision boundary

Given a linearly separable D, a linear decision boundary separating negatives from positives can be obtained using, for instance, **PLA** or **logistic regression**

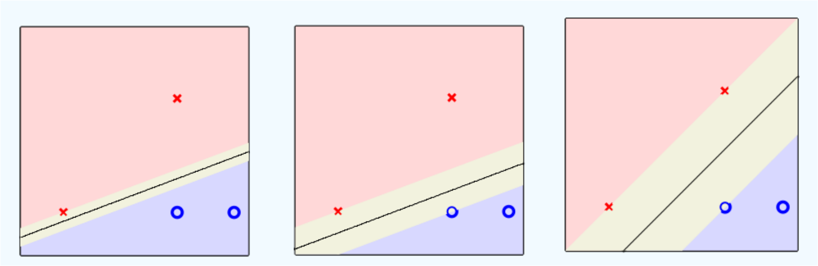


Is there one that is preferable than others ?

Intuition



Maximum margin



Any of these lines separate the negatives from the positives
They have margins of different sizes

Problem formulation

How to find the **hyperplane that maximizes the margin** ?

In **SVM**, this is achieved by formulating the problem as a quadratic programmin (QP) optimization problem

QP: optimization of quadratic functions with linear constraints on the variables

Notations

Previous Chapters

$\mathbf{x} \in \{1\} \times \mathbb{R}^d; \ \mathbf{w} \in \mathbb{R}^{d+1}$

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}; \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}.$$

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^{\mathrm{T}}\mathbf{x})$$

This Chapter

$\mathbf{x} \in \mathbb{R}^d; \ b \in \mathbb{R}, \ \mathbf{w} \in \mathbb{R}^d$

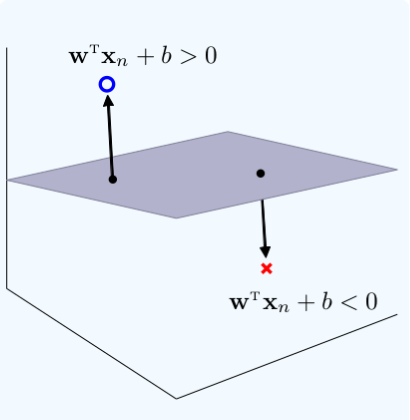
$$b = \text{bias}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}; \quad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix}.$$

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b)$$

$h(\mathbf{x}) = \mathbf{w}^T\mathbf{x}_n + b = 0$ defines a hyperplane H

Classification



Output class: $h(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^T\mathbf{x} + b)$

Relate parameters to margin

The classifier has parameters $(\mathbf{w}, b)$:

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$$

We need to somehow relate $\mathbf{w}$ and b with the margin

Margin is the distance between H and the closest point among all points in D

$\Longrightarrow$ Let us examine $d(\mathbf{x}, H)$!

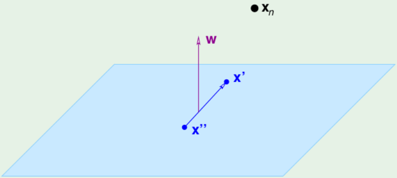
Recap: vector normal to the hyperplane

$$h(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

The vector $\mathbf{w}$ is $\perp$ to the plane in the $\mathcal{X}$ space:

Take $\mathbf{x}'$ and $\mathbf{x}''$ on the plane

??



Recap: distance between point and hyperplane

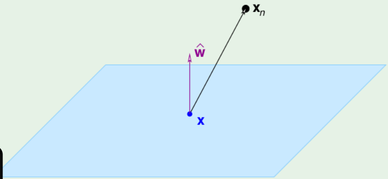
$d(\mathbf{x}_n, H) =?$

Distance between $\mathbf{x}_n$ and the plane:

Take any point $\mathbf{x}$ on the plane

Projection of $\mathbf{x}_n - \mathbf{x}$ on $\mathbf{w}$

??



Recap: distance between point and hyperplane

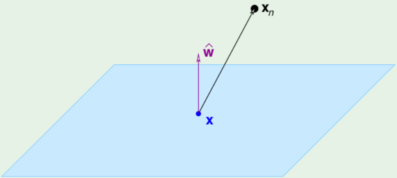
$d(\mathbf{x}_n, H) = ?$

Distance between $\mathbf{x}_n$ and the plane:

Take any point $\mathbf{x}$ on the plane

Projection of $\mathbf{x}_n - \mathbf{x}$ on $\mathbf{w}$

$\hat{\mathbf{w}} = \dfrac{\mathbf{w}}{\|\mathbf{w}\|} \implies \text{distance} = \left|\hat{\mathbf{w}}^\intercal(\mathbf{x}_n - \mathbf{x})\right|$



Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})|$$

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

$$\mathbf{w}^T(\mathbf{x_n} - \mathbf{x}) = \mathbf{w}^T\mathbf{x_n} - \mathbf{w}^T\mathbf{x}$$

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

$$\mathbf{w}^T(\mathbf{x_n} - \mathbf{x}) = \mathbf{w}^T\mathbf{x_n} - \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{x_n} + b - (\mathbf{w}^T\mathbf{x} + b)$$

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

$$\mathbf{w}^T(\mathbf{x_n} - \mathbf{x}) = \mathbf{w}^T\mathbf{x_n} - \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{x_n} + b - (\mathbf{w}^T\mathbf{x} + b)$$

$$= \mathbf{w}^T\mathbf{x_n} + b - 0 = \mathbf{w}^T\mathbf{x_n} + b$$

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

$$\mathbf{w}^T(\mathbf{x_n} - \mathbf{x}) = \mathbf{w}^T\mathbf{x_n} - \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{x_n} + b - (\mathbf{w}^T\mathbf{x} + b)$$

$$= \mathbf{w}^T\mathbf{x_n} + b - 0 = \mathbf{w}^T\mathbf{x_n} + b$$

Why $\mathbf{w}^T\mathbf{x} + b = 0$ **?**

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

$$\mathbf{w}^T(\mathbf{x_n} - \mathbf{x}) = \mathbf{w}^T\mathbf{x_n} - \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{x_n} {\color{red}+b-}(\mathbf{w}^T\mathbf{x} + {\color{red}b})$$

$$= \mathbf{w}^T\mathbf{x_n} + b - 0 = \mathbf{w}^T\mathbf{x_n} + b$$

$$dist(\mathbf{x}_n, H) = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T\mathbf{x_n} + b| = \frac{1}{||\mathbf{w}||}y_n(\mathbf{w}^T\mathbf{x_n} + b)$$

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

$\mathbf{w}^T$

$= \mathbf{w}$

Why I can do $|\mathbf{w}^T\mathbf{x_n} + b| = y_n\left(\mathbf{w}^T\mathbf{x}_n + b\right)$?

$$dist(\mathbf{x}_n, H) = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T\mathbf{x_n} + b| = \frac{1}{||\mathbf{w}||}y_n(\mathbf{w}^T\mathbf{x_n} + b)$$

Rewriting the distance between point and hyperplane

$$dist(\mathbf{x}_n, H) = |\hat{\mathbf{w}}^T(\mathbf{x_n} - \mathbf{x})| = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T(\mathbf{x_n} - \mathbf{x})|$$

$$\mathbf{w}^T(\mathbf{x_n} - \mathbf{x}) = \mathbf{w}^T\mathbf{x_n} - \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{x_n} + b - (\mathbf{w}^T\mathbf{x} + b)$$

$$= \mathbf{w}^T\mathbf{x_n} + b - 0 = \mathbf{w}^T\mathbf{x_n} + b$$

$$dist(\mathbf{x}_n, H) = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T\mathbf{x_n} + b| = \frac{1}{||\mathbf{w}||}y_n(\mathbf{w}^T\mathbf{x_n} + b)$$

(because if $\mathbf{x_n}$ is at the correct side $\implies y_n(\mathbf{w}^T\mathbf{x_n}) > 0$)

Choosing a convenient hyperplane representation (weights)

Distance as seen before:

$$dist(\mathbf{x}_n, H) = \frac{1}{||\mathbf{w}||}|\mathbf{w}^T\mathbf{x_n} + b| = \frac{1}{||\mathbf{w}||}y_n(\mathbf{w}^T\mathbf{x_n} + b)$$

If I manage to make $|\mathbf{w}^T\mathbf{x}_n + b| = 1$, then I will have

$$dist(\mathbf{x}_n, H) = \frac{1}{||\mathbf{w}||}$$

We can always rescale $(\mathbf{w}, b)$ so as to have $y_n\left(\mathbf{w}^T\mathbf{x}_n + b\right) = 1$.

Let us do that with respect to the closest point to the hyperplane:

$$\rho = \min_{n=1,\ldots,N} \; y_n(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n + b),$$

If we divide $(\mathbf{w}, b)$ by ρ, the hyperplane does not change:

$$\min_{n=1,\ldots,N} y_n\left(\frac{\mathbf{w}}{\rho}^{\mathrm{T}}\mathbf{x}_n + \frac{b}{\rho}\right) = \frac{1}{\rho}\min_{n=1,\ldots,N} \; y_n(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n + b) = \frac{\rho}{\rho} = 1.$$

Homework

Exercise 8.2

Consider the data below and a 'hyperplane' $(b, \mathbf{w})$ that separates the data.

$$\mathrm{X} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 1.2 \\ -3.2 \end{bmatrix} \quad b = -0.5$$

(a) Compute $\rho = \min_{n=1,\ldots,N} y_n(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n + b)$.

(b) Compute the weights $\frac{1}{\rho}(b, \mathbf{w})$ and show that they satisfy (8.2).

(c) Plot both hyperplanes to show that they are the *same* separator.

(8.2) $$\min_{n=1,\ldots,N} y_n(\mathbf{w}^T\mathbf{x}_n + b) = 1$$

Wrapping up

Distance between $\mathbf{x}_n$ and the hyperplane H defined by $(\mathbf{w}, b)$:

$$dist(\mathbf{x}_n, H) = \frac{1}{||\mathbf{w}||} y_n(\mathbf{w}^T \mathbf{x_n} + b)$$

We can always choose $(\mathbf{w}, b)$ such that the closest point $\mathbf{x}_n$ to H satisfies

$$y_n(\mathbf{w}^T \mathbf{x_n} + b) = 1$$

In such case

$$dist(\mathbf{x}_n, H) = \frac{1}{||\mathbf{w}||}$$

Thus, note that H separates D iff it can be represented by $(\mathbf{w}, b)$ such that

$$\min_{n=1,\ldots,N} y_n(\mathbf{w}^T\mathbf{x}_n + b) = 1$$

The problem we want to solve

$$\begin{aligned} \text{maximize} \quad & \frac{1}{||\mathbf{w}||} \\ \text{subject to} \quad & \min_{i=1,\ldots,N} y_i(\mathbf{w}^T\mathbf{x}_i + b) = 1 \end{aligned}$$

The problem we want to solve

$$\text{maximize} \quad \frac{1}{||\mathbf{w}||}$$
$$\text{subject to} \quad \min_{i=1,\ldots,N} y_i(\mathbf{w}^T\mathbf{x}_i + b) = 1$$

The constraint $\min_{i=1,\ldots,N} y_i(\mathbf{w}^T\mathbf{x}_i + b) = 1$ implies $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$ which has the effect of forcing all examples to be classified correctly

The equality $\min_{i=1,\ldots,N} y_i(\mathbf{w}^T\mathbf{x}_i + b) = 1$ implies that the distance of the closest point to the hyperplane is $\frac{1}{||\mathbf{w}||}$ (a nice objective function!)

The problem we want to solve

$$\begin{aligned} \text{maximize} \quad & \frac{1}{||\mathbf{w}||} \\ \text{subject to} \quad & \min_{i=1,\ldots,N} y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1 \end{aligned}$$

Equivalent formulation

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{w}^T \mathbf{w} \\ \text{subject to} \quad & \min_{i=1,\ldots,N} y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1 \end{aligned}$$

Relaxed formulation

Original minimization formulation:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{subject to} \quad & \min_{i=1,\dots,N} y_i(\mathbf{w}^T\mathbf{x}_i + b) = 1 \end{aligned}$$

Equivalent relaxed formulation:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{subject to} \quad & y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, i = 1, \dots, N \end{aligned}$$

The equivalence can be proved by contradiction (see Chapter on SVM, page 7)

A toy example

Contraints ??: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$?

$$X = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix}$$

A toy example

$$X = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix}$$

Contraints ??: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$?

$$\begin{aligned} -b &\geq 1 \quad (1) \\ -(2w_1 + 2w_2 + b) &\geq 1 \quad (2) \\ 2w_1 + b &\geq 1 \quad (3) \\ 3w_1 + b &\geq 1 \quad (4) \end{aligned}$$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

Example: Solving it by hand

$$\begin{aligned} -b &\geq 1 \quad (1) \\ -(2w_1 + 2w_2 + b) &\geq 1 \quad (2) \\ 2w_1 + b &\geq 1 \quad (3) \\ 3w_1 + b &\geq 1 \quad (4) \end{aligned}$$

- From (3) and (1)
  $2w_1 + b \geq 1$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1 - b)$

Example: Solving it by hand

$$\begin{aligned} -b &\geq 1 \quad (1)\\ -(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\ 2w_1 + b &\geq 1 \quad (3)\\ 3w_1 + b &\geq 1 \quad (4) \end{aligned}$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1-b)$ && $b \leq -1$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1-b)$ && $b \leq -1$
  $\Longrightarrow w_1 \geq 1$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1 - b)$ && $b \leq -1$
  $\Longrightarrow w_1 \geq 1$

- From (2) and (3):
  $-(2w_1 + 2w_2 + b) \geq 1$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

- From (3) and (1)
  $2w_1 + b \geq 1 \leadsto 2w_1 \geq 1 - b \leadsto w_1 \geq \frac{1}{2}(1 - b)$ && $b \leq -1$
  $\implies w_1 \geq 1$

- From (2) and (3):
  $-(2w_1 + 2w_2 + b) \geq 1 \leadsto -2w_1 - 2w_2 - b \geq 1$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1 - b)$ && $b \leq -1$
  $\Longrightarrow w_1 \geq 1$

- From (2) and (3):
  $-(2w_1 + 2w_2 + b) \geq 1 \rightsquigarrow -2w_1 - 2w_2 - b \geq 1 \rightsquigarrow$
  $2w_2 \leq -2w_1 - b - 1$

Example: Solving it by hand

$$\begin{aligned} -b &\geq 1 \quad (1)\\ -(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\ 2w_1 + b &\geq 1 \quad (3)\\ 3w_1 + b &\geq 1 \quad (4) \end{aligned}$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1 - b)$ && $b \leq -1$
  $\Longrightarrow w_1 \geq 1$

- From (2) and (3):
  $-(2w_1 + 2w_2 + b) \geq 1 \rightsquigarrow -2w_1 - 2w_2 - b \geq 1 \rightsquigarrow$
  $2w_2 \leq -2w_1 - b - 1$ && $2w_1 + b \geq 1$

Example: Solving it by hand

$$
\begin{aligned}
-b &\geq 1 \quad (1)\\
-(2w_1 + 2w_2 + b) &\geq 1 \quad (2)\\
2w_1 + b &\geq 1 \quad (3)\\
3w_1 + b &\geq 1 \quad (4)
\end{aligned}
$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1 - b)$ && $b \leq -1$
  $\Longrightarrow w_1 \geq 1$

- From (2) and (3):
  $-(2w_1 + 2w_2 + b) \geq 1 \rightsquigarrow -2w_1 - 2w_2 - b \geq 1 \rightsquigarrow$
  $2w_2 \leq -2w_1 - b - 1$ && $2w_1 + b \geq 1 \Longrightarrow w_2 \leq -1$

Example: Solving it by hand

$$-b \geq 1 \quad (1)$$
$$-(2w_1 + 2w_2 + b) \geq 1 \quad (2)$$
$$2w_1 + b \geq 1 \quad (3)$$
$$3w_1 + b \geq 1 \quad (4)$$

- From (3) and (1)
  $2w_1 + b \geq 1 \rightsquigarrow 2w_1 \geq 1 - b \rightsquigarrow w_1 \geq \frac{1}{2}(1-b)$ && $b \leq -1$
  $\implies w_1 \geq 1$

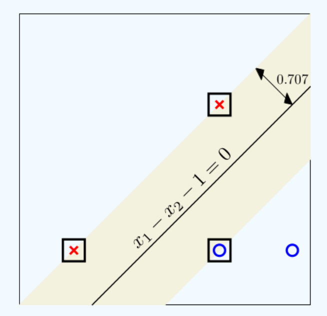
- From (2) and (3):
  $-(2w_1 + 2w_2 + b) \geq 1 \rightsquigarrow -2w_1 - 2w_2 - b \geq 1 \rightsquigarrow$
  $2w_2 \leq -2w_1 - b - 1$ && $2w_1 + b \geq 1 \implies w_2 \leq -1$

Thus, $\frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}(w_1^2 + w_2^2) \geq 1$ and the minimum is at $\mathbf{w} = (1, -1)$;
$(b = -1, w_1 = 1, w_2 = -1)$ satisfies the 4 constraints

Example: the solution

The separating hyperplane H with maximum margin is given by $x_1 - x_2 - 1 = 0$.

$$X = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix}$$



The margin is $\frac{1}{||\mathbf{w}||} = \frac{1}{\sqrt{2}} \approx 0.707$

Summary

- (linearly separable case) The goal is to find a hyperplane that maximizes the margin
- We examined the formulation of the hard margin SVM
- It can be written as a QP otimization (quadratic objective function and linear inequality constraints)
- We solved a toy example by hand
- In practice, we use *solvers*