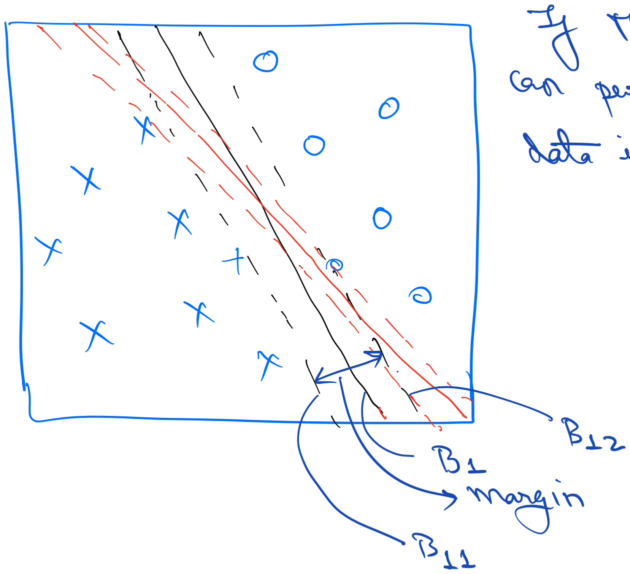
Support vector Machines



If there is a hyperplane that can perfectly classify data $\Longrightarrow$ data is linearly separable

We want to find a boundary that is a maximum margin hyperplane

→ A small margin ⇒ boundary highly sensitive to noise or slightly different dataset. Hence worse generalization on unseen dataset.

→ Bigger margin ⟶ less complex model & hence better generalization

⇒ For a Binary class problem

labels $y_i \in \{-1, 1\}$

Let $w^T x + b = 0$: equation of separating hyperplane ($d-1$ dimension) that separates the two classes.

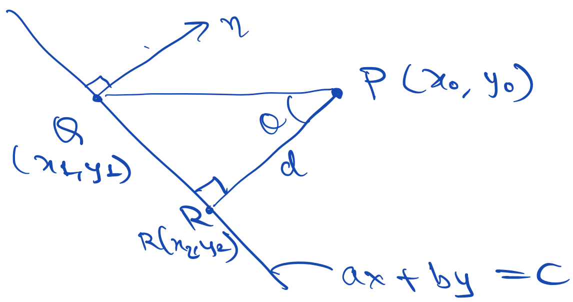
In 2d : It's a line.
In 3d : It's a plane
⋮

This is achieved by placing them on opposite sides. Hence:

$$w^T x_i + b > 0 \qquad y_i = 1$$

$$w^T x_i + b < 0 \qquad y_i = -1$$

Distance of a point from a line



$$d = \|QP\| \cos\theta$$

$$= \frac{\|QP\|\,\|n\| \cos\theta}{\|n\|}$$

$$= \frac{\langle QP \cdot n \rangle}{\|n\|}$$

$$= \frac{|(x_0 - x_1, y_0 - y_1) \cdot (a \cdot b)|}{\sqrt{a^2 + b^2}}$$

$$= \frac{|a(x_0 - x_1) + b(y_0 - y_1)|}{\sqrt{a^2 + b^2}}$$

$$= \frac{|ax_0 - ax_1 + by_0 - by_1|}{\sqrt{a^2 + b^2}}$$

$$= \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$

vector orthogonal to $ax + by + c = 0$

$Q: ax_1 + by_1 = -c$ — (i)

$R: ax_2 + by_2 = -c$ — (ii)

(i) — (ii)

$$a(x_1 - x_2) + b(y_1 - y_2) = 0$$

$$(a, b) \cdot \underbrace{(x_1 - x_2,\ y_1 - y_2)}_{\overrightarrow{RQ} \text{ along direction of line}} = 0$$

$\Rightarrow (a, b)$ is a vector orthogonal to the line

$n : (a, b)$

$\Rightarrow \quad D(x) = \dfrac{|w^T x + b|}{\|w\|}$ $\left\{ \text{distance of any point in the feature space to the hyperplane} \right.$

Let

$K_+ > 0$: Distance of closest pt with $y = 1$ from hyperplane.

$K_- > 0$: Distance of closest pt with $y = -1$ from hyperplane

$\Rightarrow$ we have constraints

$$\frac{w^T x_i + b}{\|w\|} \geq K_+ \quad \text{if } y_i = 1$$

$$\frac{w^T x_i + b}{\|w\|} \leq -K_- \quad \text{if } y_i = -1$$

$\Rightarrow$ combining the 2 equations & let $M = K_+ = K_-$

$\Rightarrow$ Margin $= 2M$ $\{$ M distance from either side of the line $\}$

$\Rightarrow$ We want to find maximum margin hyperplane that follows the above constraints:

$$\max_{w, b} M \qquad \text{---} (4.72)$$

$$\text{subject to: } y_i(w^T x_i + b) \geq M \|w\|$$

If any w, b satisfy above constraints then their scaled versions would also satisfy it.

→ Conviniently choosing $\|w\| = \frac{1}{M}$ so $\max M \iff \min \|w\|$

→ we have equivalent problem:

$$\min \|w\|$$

or

$$\boxed{\begin{aligned} &\min_{w,b} \frac{\|w\|^2}{2} \\ &\text{st: } \quad y_i(w^T x_i + b) \geq 1 \end{aligned}}$$

→ Quadratic Programming Problem