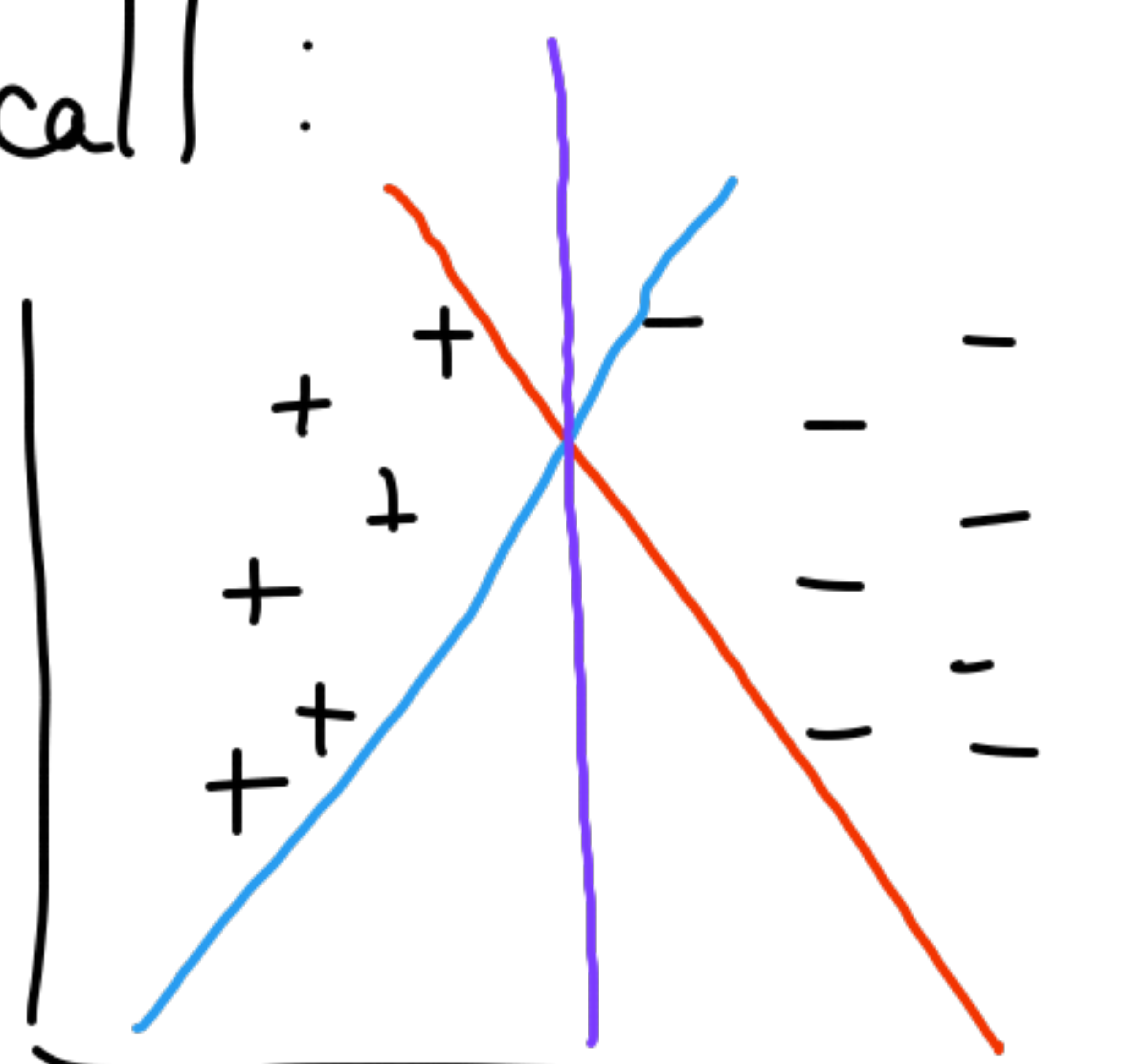
CS725

SUPPORT VECTOR MACHINES (SVMs)

Recall:



Which of these hyperplanes would you prefer? PURPLE ONE!

[Perceptron can learn any of the hyperplanes in the figure.]

There's a unique hyperplane that maximizes the margin; SVMs identify this hyperplane

Recap: The distance of a point x from a hyperplane $\mathcal{H}: w^T x + b = 0$

$$\Rightarrow \frac{|w^T x + b|}{\|w\|_2}$$

Margin of $\mathcal{H}$ w.r.t. a dataset $\mathcal{D}$: $\gamma = \min_{x \in \mathcal{D}} \frac{|w^T x + b|}{\|w\|_2}$

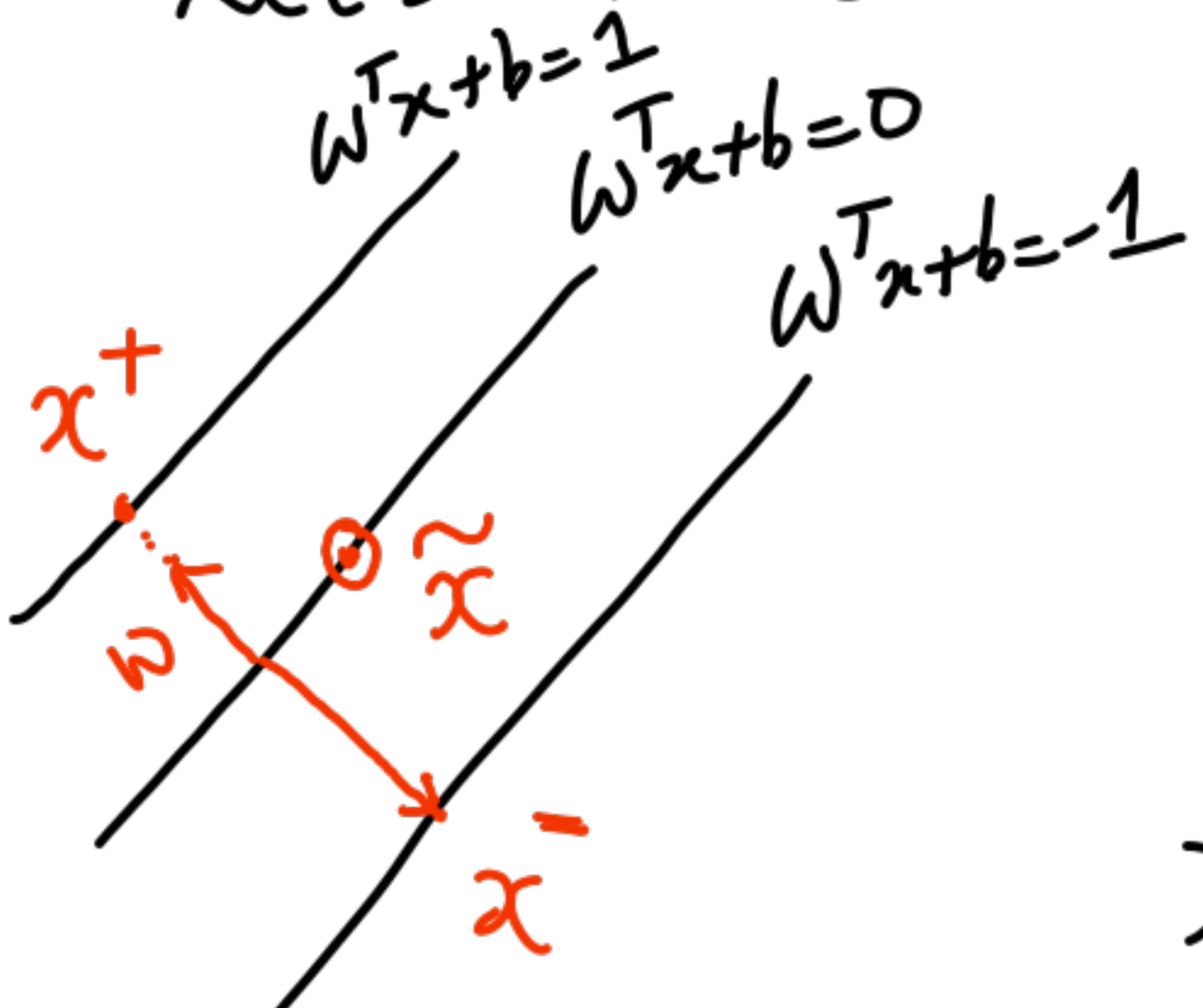
Property of margins: They are scale invariant.

$$\min_x \frac{|w^T x + b|}{\|w\|_2} = \min_x \frac{|\alpha w^T x + \alpha b|}{\|\alpha w\|_2}, \alpha \neq 0$$

Without loss of generality, let us scale w and b such that

$$\min_{x} |w^T x + b| = 1 \Rightarrow \gamma = \frac{1}{\|w\|_2}$$

Let's derive the expression for γ again,



$$x^+ = \tilde{x} + \gamma \frac{w}{\|w\|}$$

$$x^- = \tilde{x} - \gamma \frac{w}{\|w\|}$$

$$\underbrace{x^+ = x^- + 2\gamma \frac{w}{\|w\|}}_{A} \; ; \; \underbrace{w^T x^+ + b = 1}_{B}$$

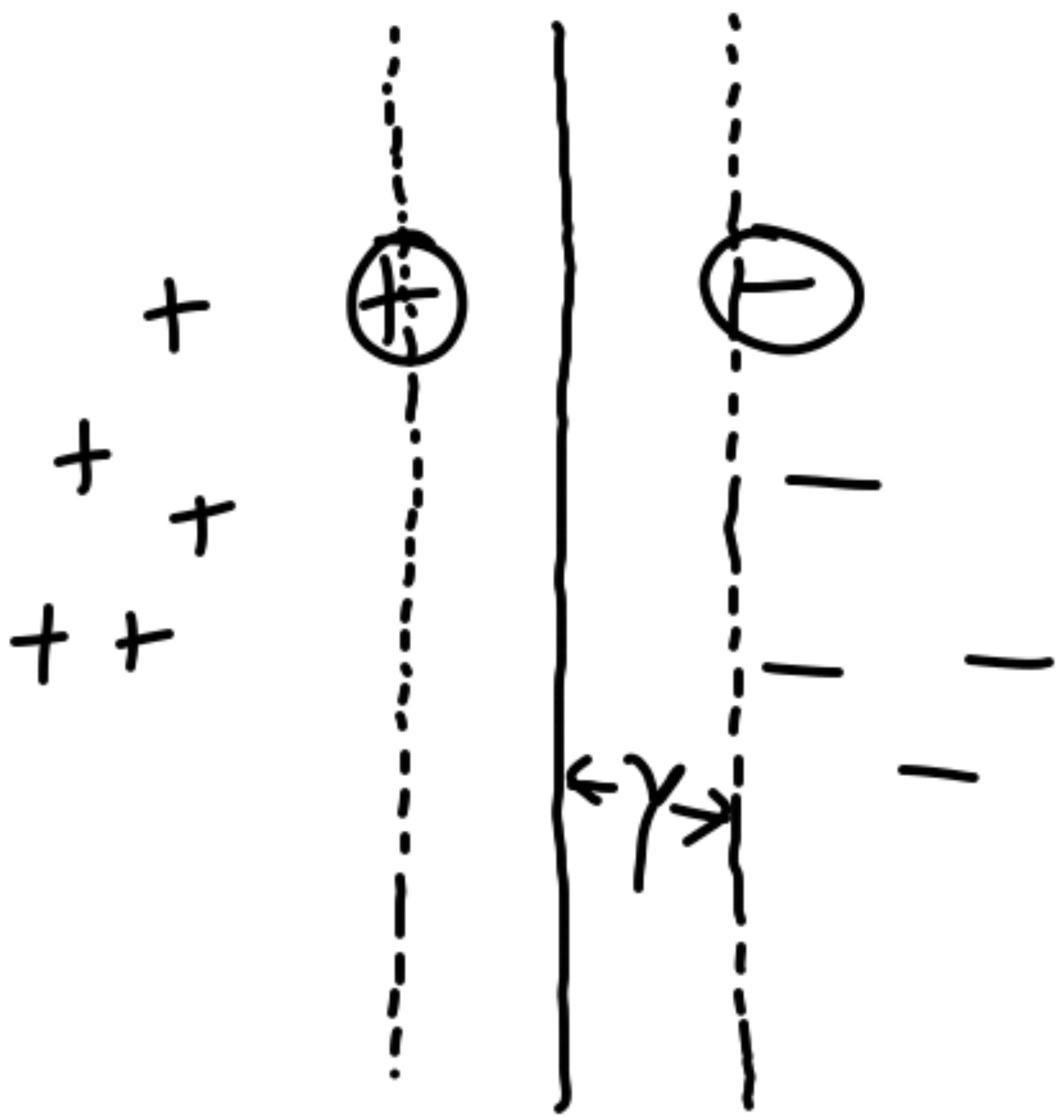
Substituting A in B,

$$w^T\left(x^- + 2\gamma \frac{w}{\|w\|}\right) + b = 1$$

$$2\gamma \frac{w^T w}{\|w\|} = 2$$

$$\Rightarrow \boxed{\gamma = \frac{1}{\|w\|_2}}$$

SVMs aim to maximize twice the margin while correctly classifying all the training points.



$$\max_{w,b} \frac{2}{\|w\|_2} \equiv \boxed{\min_{w,b} \frac{1}{2}\|w\|^2}$$

maximize margin

$$\text{s.t.} \quad \boxed{y_i\left(w^T x_i + b\right) \geq 1, \; i \in \{1, \dots, n\}}$$

$$x_i \in \mathbb{R}^d, \; y_i \in \{-1, 1\}$$

all training pts are correctly classified

$$y_i\left(w^T x_i + b\right) \geq 0$$

$$\min_x \left|w^T x + b\right| = 1$$

$$\Updownarrow$$

$$y_i\left(w^T x_i + b\right) \geq 1$$

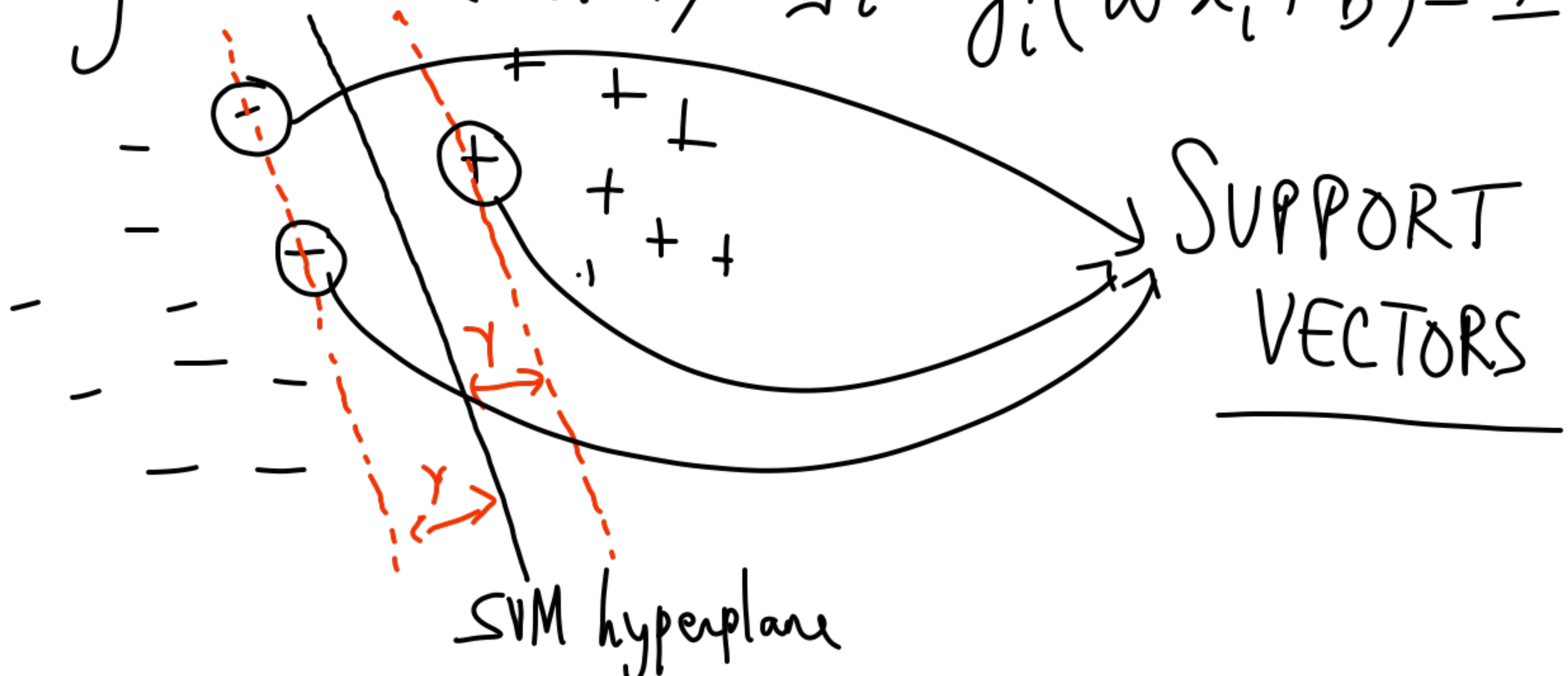
HARD-MARGIN SVM CLASSIFIER

$$\boxed{\begin{aligned} &\min_{w,b} \frac{1}{2}\|w\|^2 \\ &\text{s.t. } \forall i \text{ in } \mathcal{D},\ y_i(w^T x_i + b) \geq 1 \end{aligned}}$$

The hard-margin SVM is a quadratic program with linear constraints

Hard-margin SVMs $\Rightarrow$ SVM is very strict about classifying training pts correctly

The optimal solution of the hard margin SVM will have some points satisfying the boundary condition i.e., $\exists i \quad y_i(w^T x_i + b) = \underline{1}$



In SVMs, support vectors fully determine the decision boundary. Removing pts from $\mathcal{D}$ that are not support vectors does not change the solution.

Hard-margin SVMs overly focus on perfectly separating the training points $\Rightarrow$ What if some pts violate the margin constraints?

Consider some points that do not satisfy the margin constraint

$$\exists i \quad y_i(w^T x_i + b) < 1 \Rightarrow 1 - y_i(w^T x_i + b) > 0$$

$$\text{Constraint violation} = \begin{cases} 0 & \text{if } 1 - y_i(w^T x_i + b) \leq 0 \\ 1 - y_i(w^T x_i + b) & \text{otherwise} \end{cases}$$

Minimize constraint violations over the complete dataset ; Type of hinge loss

$$\mathcal{L}_{s\text{-}SVM} = \sum_i \max\left[0, 1 - y_i(w^T x_i + b)\right]$$

Along with maximizing the margin, we formulate the following unconstrained optimization problem

$$\underbrace{\min_{w,b} \frac{1}{2}\|w\|^2 + C \cdot \sum_i \max\left[0, 1 - y_i(w^T x_i + b)\right]}_{\text{SOFT-MARGIN SVM FORMULATION}}$$

C is a hyperparameter

In the soft-SVM objective, C is the trade-off between fit to training points and maximizing the margin.

If C is small, SVM is relaxed/loose, larger margins, more forgiving of constraint violations

An alternate view of the soft-SVM objective is as a REGULARIZED HINGE LOSS;

$$\underbrace{\min_{w,b} \frac{1}{2}\|w\|^2}_{\text{Regularizer on } w} + \underbrace{C\sum_i \max\left(0, 1-y_i\left(w^T x_i + b\right)\right)}_{\text{fit to training data}}$$

An equivalent constrained optimization problem for the soft-margin SVM:

$$\boxed{\begin{aligned}&\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{s.t. } y_i(w^T x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0\ \forall i\end{aligned}}$$

ξ_i's are referred to as SLACK VARIABLES

ξ_i's are capturing the allowed amount of "SLACK" when solving for the soft-margin SVM