# Video 11

Saturday, October 10, 2020 5:00 PM

**Binary classification task**

(using same notation than for perceptron)

inputs are associated with targets $\in \{-1, 1\}$

negative class — positive class

Linear model : $z = \vec{w}^T \vec{x} + b$

$$y = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases} = \text{sign}(z)$$

Error / loss / cost of our model:

$$\mathcal{L}_{0-1}(y, t) = \begin{cases} 0 & \text{if } y = t \\ 1 & \text{if } y \neq t \end{cases}$$

$y$: model predictions, $t$: targets

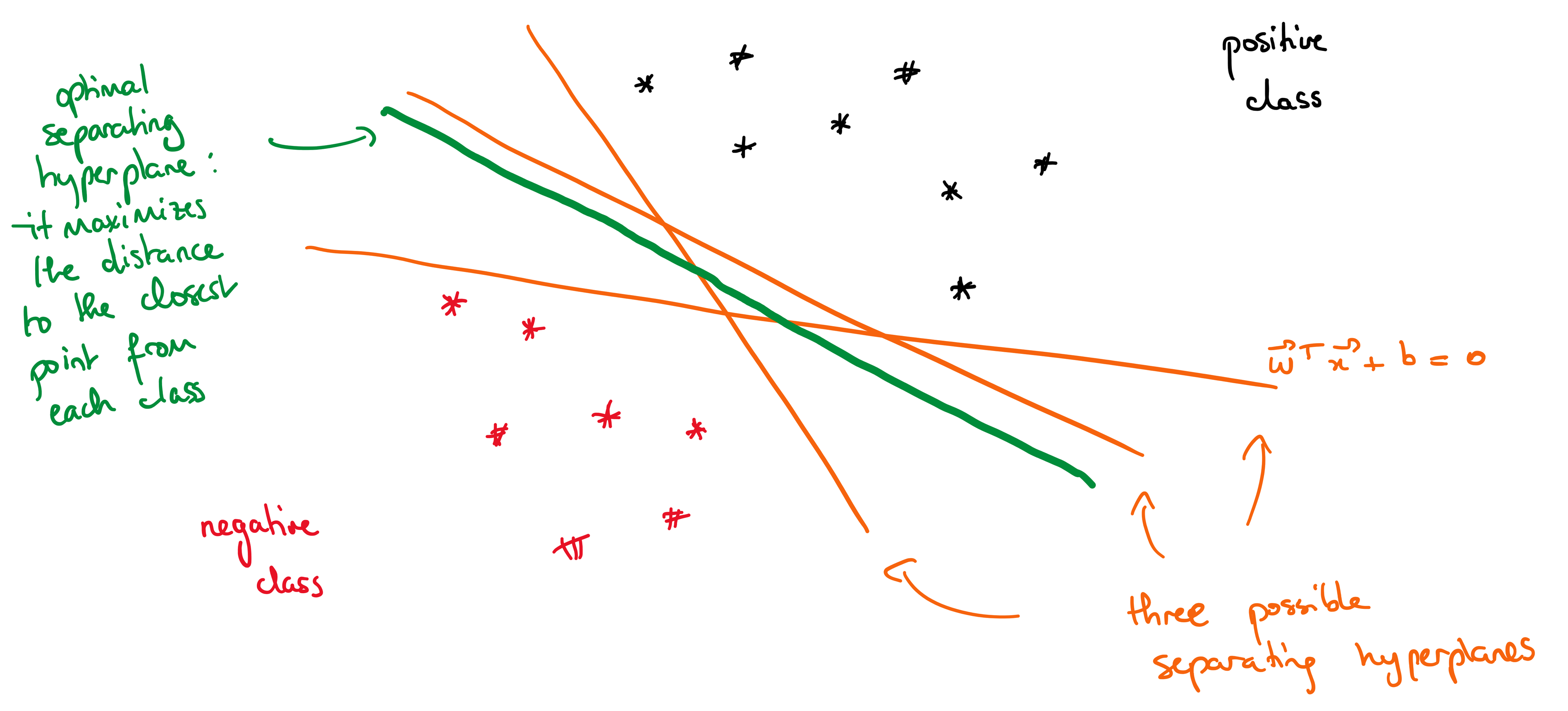
$z = \vec{w}^T x + b$ , if we replace $\vec{w} \leftarrow 2\vec{w}$, $b \leftarrow 2b$

$z$ gets modified but $y$ is identical.

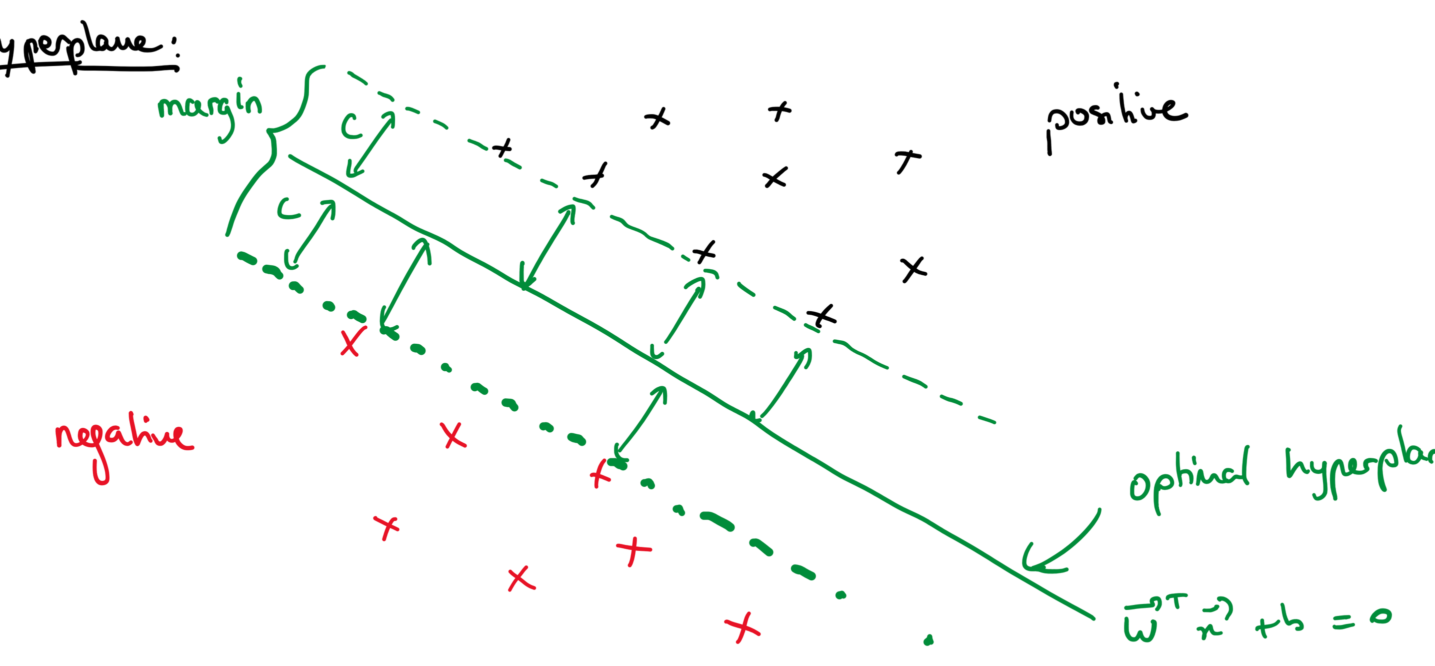
$$\left.\begin{aligned} y &= \text{sign}(z) \\ y &= \text{sign}(2z) \end{aligned}\right\} \text{equal}$$

$\Rightarrow$ The classifier's confidence is not captured by this loss.

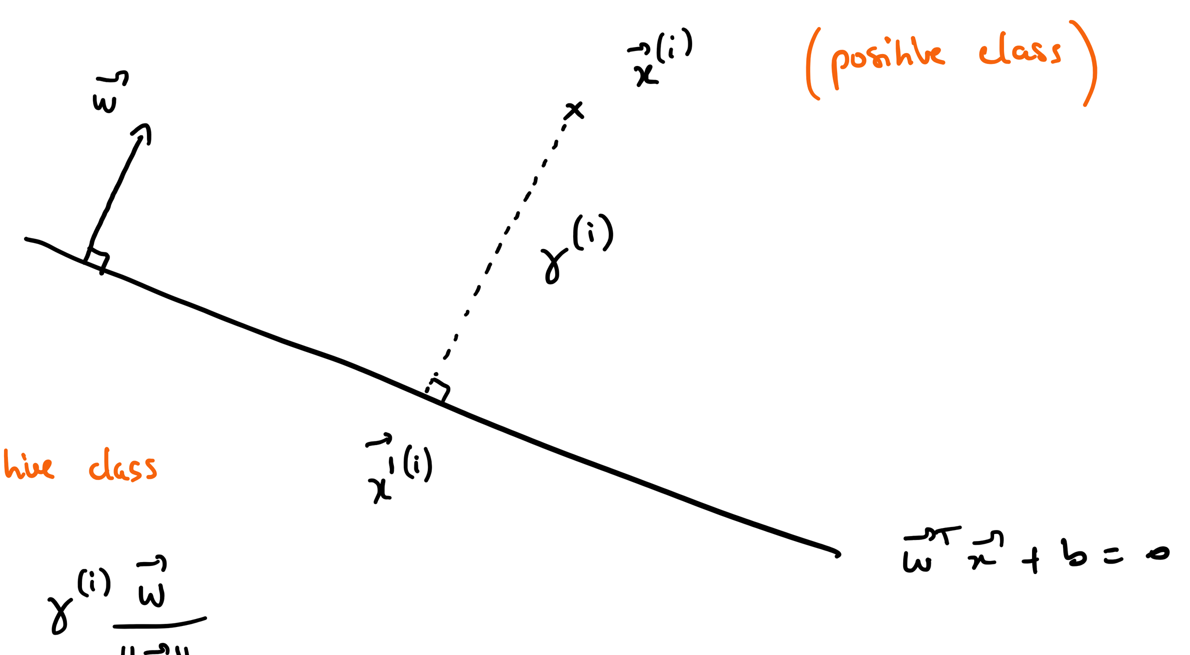
Separating hyperplanes




**Max-margin hyperplane:**




**Geometry reminders**




For positive class

$$\vec{x}'^{(i)} = \vec{x}^{(i)} - \gamma^{(i)} \underbrace{\frac{\vec{w}}{\|\vec{w}\|}}_{\text{unit vector}}$$

(if negative class − becomes +)

How to measure $\gamma^{(i)}$:

Given that $\vec{x}'^{(i)}$ is on the hyperplane : $\vec{w}^T \vec{x}'^{(i)} + b = 0$

$$\vec{w}^T \left( \vec{x}^{(i)} - \gamma^{(i)} \frac{\vec{w}}{\|\vec{w}\|} \right) + b = 0$$

$\vec{w}^T \vec{w} = \|w\|^2$

$$\vec{w}^T \vec{x}^{(i)} - \gamma^{(i)} \frac{\vec{w}^T \vec{w}}{\|\vec{w}\|} + b = 0$$

$$\vec{w}^T \vec{x}^{(i)} - \gamma^{(i)} \|\vec{w}\| + b = 0$$

$$\boxed{\gamma^{(i)} = \frac{\vec{w}^T \vec{x}^{(i)} + b}{\|\vec{w}\|}}$$

(this holds when the example is in the positive class)

The following holds for both positive and negative examples.

$$\gamma^{(i)} = t^{(i)} \left( \frac{\vec{w}^T \vec{x}^{(i)} + b}{\|\vec{w}\|} \right)$$

**Enforcing the geometric margin:**

For an example $\vec{x}^{(i)}$, $\gamma^{(i)} \geq C$

$C$: minimum distance between the training point $\vec{x}^{(i)}$ and the classifier's decision boundary (the hyperplane)

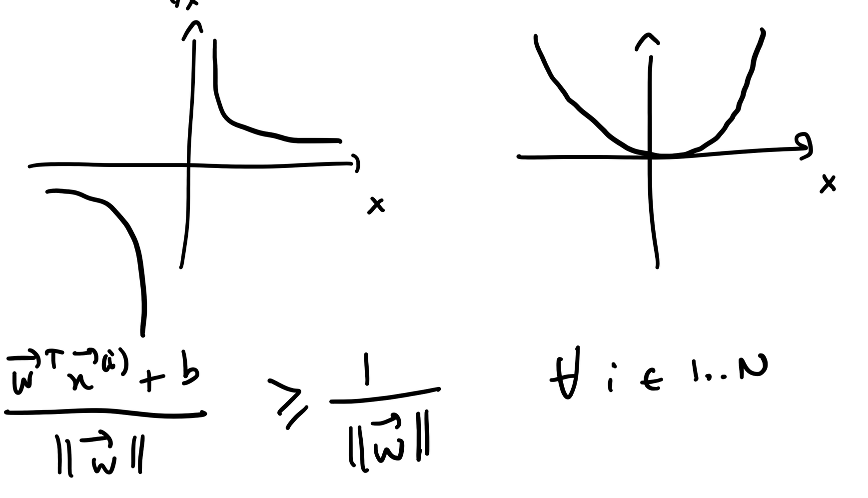
We want a classifier such that:

$$\boxed{\max_{\vec{w}, b} C \quad \text{s.t.} \quad t^{(i)} \frac{\vec{w}^T \vec{x}^{(i)} + b}{\|\vec{w}\|} \geq C \quad \forall i \in 1..N}$$

$N$: nb of training examples

* Set $C = \frac{1}{\|\vec{w}\|}$ (because the classifier is invariant to rescaling of $\vec{w}$)

* $\max \frac{1}{x} \Leftrightarrow \min x^2$




$$\min_{\vec{w}, b} \|\vec{w}\|^2 \quad \text{s.t.} \quad t^{(i)} \frac{\vec{w}^T \vec{x}^{(i)} + b}{\|\vec{w}\|} \geq \frac{1}{\|\vec{w}\|} \quad \forall i \in 1..N$$

$$\boxed{\min_{\vec{w}, b} \|\vec{w}\|^2 \quad \text{s.t.} \quad t^{(i)} \left( \vec{w}^T \vec{x}^{(i)} + b \right) \geq 1 \quad \forall i \in 1..N}$$