Lecture 6 | 2/20/18 | 11 650

Null model

$\vec{x}$'s

$\mathcal{H} = \{0, 1\}$

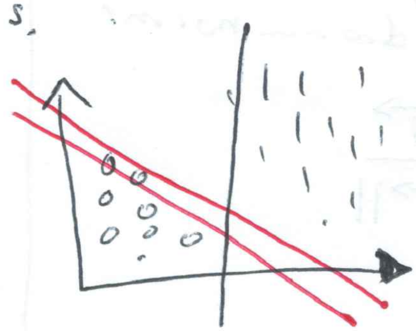
What would be $g = \mathcal{A}(\mathbb{D} = y, \mathcal{H}) = \text{Model}[\vec{y}]$

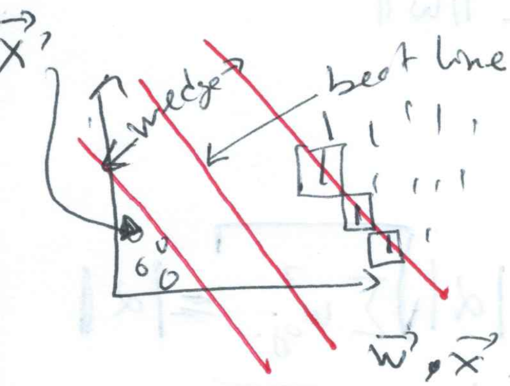
$\mathcal{H} = \{\mathbb{1}_{\vec{w}, \vec{x} > 0}, \text{ s.t } \vec{w} \in \mathbb{R}^{p+1}\}$

begins @ $\vec{1}$.

Assume linear separables.

Which hyperplane spec by $\vec{w}$ no bear?





→ "max-margin hyperplane" 1963 Vapnik

$\vec{w}, \vec{x} \longrightarrow g = \mathbb{1}_{\vec{w}\cdot\vec{x}}$ which is called the $\underbrace{\text{support Vector}}_{\text{Central observation}}$ $\underbrace{\text{machine}}_{\text{model}}$

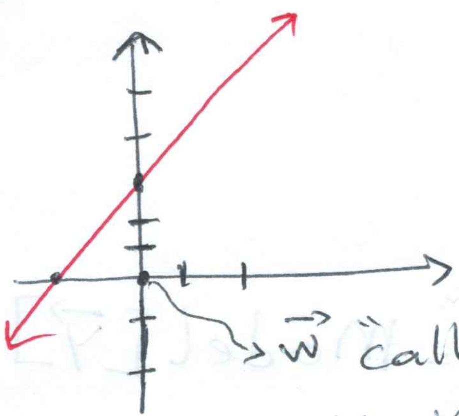
$\{\mathbb{1}_{\vec{w}\cdot\vec{x}+b>0} : \vec{w}\in\mathbb{R}^{p}, b\in\mathbb{R}\}$

x no longer has one of the final line.

$$X_2 = 2X_1 + 3 \Rightarrow 2X_1 - X_2 + 3 = 0$$

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 3 = 0$$

$$\underbrace{\vec{w}}_{\text{Hyperplane}} \cdot \vec{v} + \underbrace{b}_{\text{constant.}} = 0$$



$\vec{w}$ "called normal vector"

No that it is orthogonal to the ~~plane~~ line

Recall from Linear algebra

$$\|\vec{w}\| := \sqrt{\sum_{j=1}^{p} w_j^2}$$

"the Euclidean norms"

$$\vec{W}_0 = \frac{\vec{w}}{\|\vec{w}\|}$$

normal vector $\vec{w}$

what do you think?

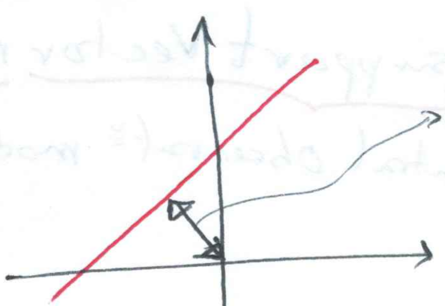
$$\|\vec{w}_0\| = \sqrt{\sum^{p} \left(\frac{w_j}{\|\vec{w}\|}\right)^2}$$

$$= \sqrt{\frac{1}{\|\vec{w}\|^2} \sum_{j=1}^{p} w_j^2} = \frac{1}{\|\vec{w}\|} \|\vec{w}\|$$

$$= 1.$$

$$\vec{l} = \alpha \vec{W}_0$$

$$\Rightarrow \|\vec{l}\| = \sqrt{\sum^{p} (\alpha \cdot w_{0j})^2} = \sqrt{\alpha^2 \sum w_{0,j}^2} = |\alpha| \underbrace{\sqrt{\sum w_{0j}^2}}_{\|\vec{w}_0\|} = |\alpha|$$



$\vec{l} = \alpha \cdot \vec{w}_0$

How to find α?

$$\Rightarrow \vec{w} \cdot \vec{l} + b = 0$$

$$\Rightarrow \vec{w} \cdot (\alpha \vec{w}_0) + b = 0 \Rightarrow \alpha \vec{w} \cdot \frac{\vec{w}}{\|\vec{w}\|} + b = 0$$

$$\Rightarrow \alpha \vec{w}$$

$$\Rightarrow \vec{w}\cdot\vec{l} + b = 0$$

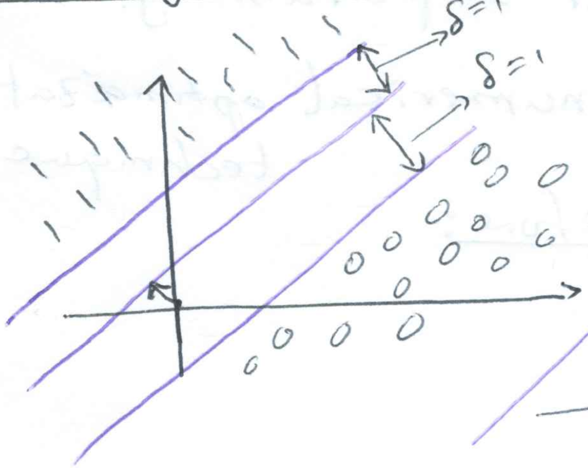
$$\Rightarrow \vec{w}\cdot\left(|\alpha|\,\vec{w_0}\right) + b = 0$$

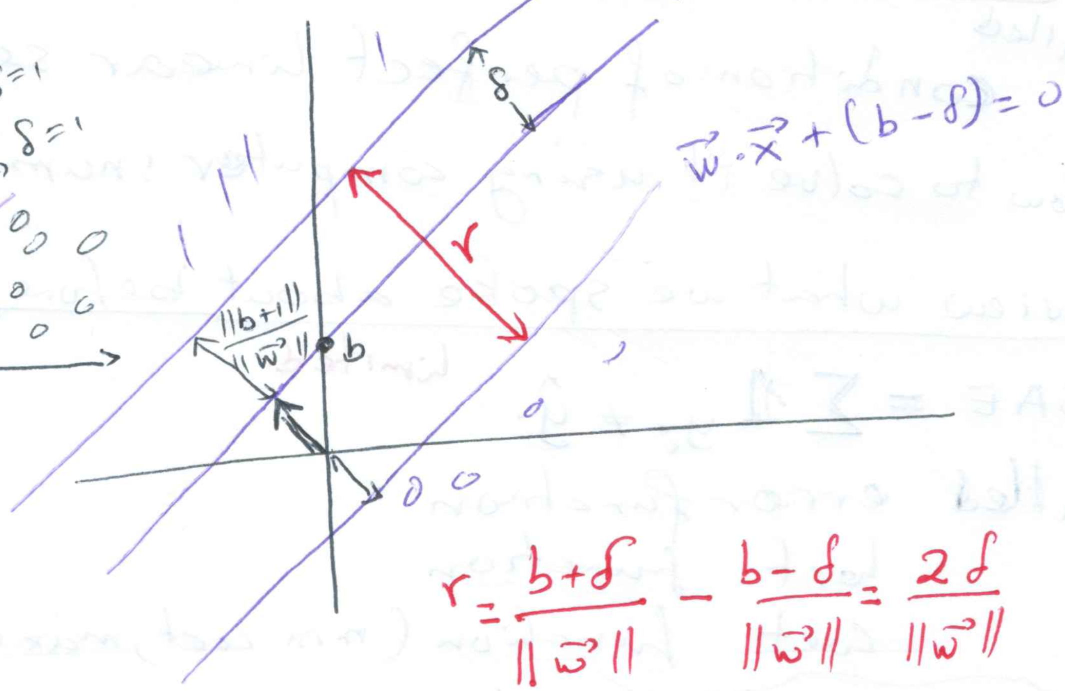
$$\Rightarrow |\alpha|\,\vec{w}\cdot\frac{\vec{w}}{\|\vec{w}\|} + b = 0$$

$$\Rightarrow |\alpha|\,\frac{\|\vec{w}\|^2}{\|\vec{w}\|} + b = 0$$

$$\boxed{\Rightarrow |\alpha| = \left|\frac{-b}{\|\vec{w}\|}\right| = \frac{1}{\|\vec{w}\|}}$$

Let's go back



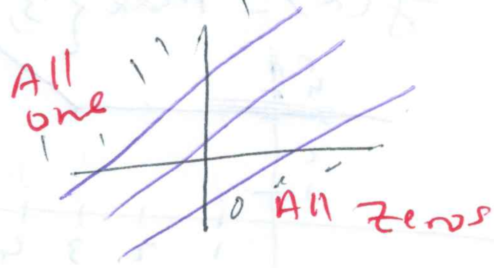


$$r = \frac{b+\delta}{\|\vec{w}\|} - \frac{b-\delta}{\|\vec{w}\|} = \frac{2\delta}{\|\vec{w}\|}$$

Coerse $\delta = 1$ means $r = \frac{b+1}{\|\vec{w}\|} - \frac{b-1}{\|\vec{w}\|} = \frac{2}{\|\vec{w}\|}$

maximizing the margin equivalently minimizing $\|\vec{w}\|$

subject to " no errors "



subject to "no errors"

$\vec{w} \cdot \vec{x} + b + 1 = 0$

$\Rightarrow \vec{w} \cdot \vec{x} + b = -1$

If $\vec{w} \cdot \vec{x} + b \geq -1 \Rightarrow y_i = 1 \ \forall_i$

If $\vec{w} \cdot \vec{x} + b \leq 1$

$$\boxed{\begin{array}{l} \text{minimize} \|\vec{w}\| \text{ s.t } \forall_i : \left(y_i - \frac{1}{2}\right)\left(\vec{w} \cdot \vec{x}_i + b\right) \geq -\frac{1}{2} \\ \text{over } \vec{w} \in \mathbb{R}^p, \ b \in \mathbb{R} \end{array}}$$

called condition of perfect linear separability.

How to solve it using computer: numerical optimization technique.

Review what we spoke about before:

limited.

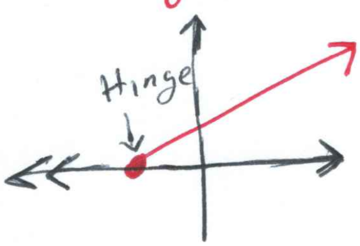
$SAE = \sum \mathbb{1}_{y_i \neq \hat{y}_i}$

called error function
loss function
cost function (min cost, max cost).

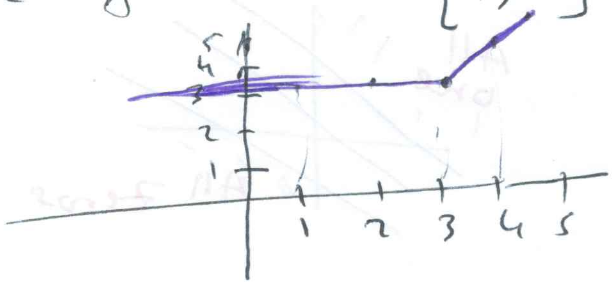
fitness function
objective function.

$H_i = \max\left\{0, -\frac{1}{2} - \left(y_i - \frac{1}{2}\right)\left(\vec{w} \cdot \vec{x}_i + b\right)\right\}$

"hinge loss"



Exple $f(x) = \max\{3, x\}$



$\{3, 4\}$ max

$\{3, 5\}$ max

$\{3, 2\}$ max

If $(y_i - \frac{1}{2})(\vec{w}\cdot\vec{x}_i + b) \geq -\frac{1}{2}$

$$(y_i - \tfrac{1}{2})(\vec{w}\cdot\vec{x}_i + b) = -\tfrac{1}{2} + d$$

$d > 0$

$$H = \min\left\{0, -\tfrac{1}{2} - \left(-\tfrac{1}{2} + d\right)\right\}$$

$$= \max\{0, -d\} = 0$$

If it is wrong we have $(y_i - \frac{1}{2})(\vec{w}\cdot\vec{x} + b) < -\frac{1}{2}$ incorrect classified

<u>Question</u>: $\min \underbrace{\frac{1}{n}\sum_{i=1}^{n} H_i}_{\text{Average hinge loss}} + \lambda \underbrace{\|\vec{u}\|^2}_{\text{maximizing the margin}}$ over $\vec{w}\in\mathbb{R}^p$, $b\in\mathbb{R}$.

what are parameters of this?

<u>Ans</u>: $\vec{w}$, b.

but λ is not a parameter

λ is called "hyperparameter."

A predefined constant. It is a tuning knob or A

what is if $\lambda \overset{\text{near}}{\approx} 0$?

~~spectrum~~ perceptrum doesn't always work.