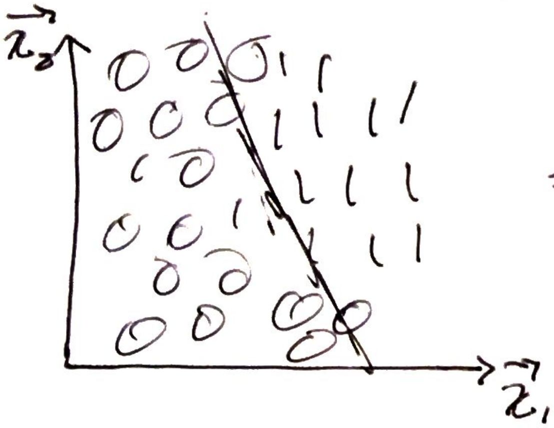
Math 342W Lecture 4

$\mathcal{H} = \{\mathbb{1}_{x_2 \geq a + bx} : a, b \in \mathbb{R}\}$ → This gives all possible linear lines



⇒ Linear Discrimination aka "Classification"

⇒ we can't find the best line, it's impossible to search all lines. (Estimation Error) No closed form s/n

$x_2 \geq a + bx$ implies $-a - bx_1 + x_2 \geq 0$, and so

$$w_0 + w_1 x_1 + w_2 x_2 \geq 0$$

w_0 ↳ bias, w_1 ↳ x_1 weight, w_2 ↳ x_2 weight

Concat 1's to all $\vec{x}$'s

$\vec{x}_1 = [1 \; x_{11} \; x_{12}]$
$\vec{x}_2 = [1 \; x_{21} \; x_{22}]$

$\mathcal{H} = \{\mathbb{1}_{\vec{w} \cdot \vec{x} \geq 0} : \vec{w} \in \mathbb{R}^3\}$ ⇒ We "over parameterized" the model

$= \mathbb{1}_{c(\vec{w} \cdot \vec{x}) \geq 0} \cdot \vec{w} \in \mathbb{R}, \forall c \in \mathbb{R}$

$$\vec{w}^* = \underset{\vec{w} \in \mathbb{R}^3}{\text{arguement}} \left\{ \sum_{i=1}^{n} \mathbb{1}_{\vec{w} \cdot \vec{x}_i \geq 0} \neq y_i \right.$$

Perceptron (1957) Algorithm

For p features:

① Initialize $\vec{w}^{t=0} = \vec{0}$ OR random $\in \mathbb{R}^{p+1}$

② Compute $\hat{y}_i = \mathbb{1}_{\vec{w}^{t=0} \cdot \vec{x}_i \geq 0}$

③ For $j = 0, 1, \ldots, p$

$$\vec{w}_0^{t=1} = \vec{w}_0^{t=0} + (y_i - \hat{y}_i)(1)$$

$$\vec{w}_1^{t=1} = \vec{w}_1^{t=0} + (y_i - \hat{y}_i)(x_{i1})$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\vec{w}_p^{t=1} = \vec{w}_p^{t=0} + \underbrace{(y_i - \hat{y}_i)}_{e_i}(x_{ip})$$

Hypothesis Space

$$\boxed{\mathcal{H} = \{\mathbb{1}_{\vec{w} \cdot \vec{x} \geq 0} : \vec{w} \in \mathbb{R}^{p+1}}$$

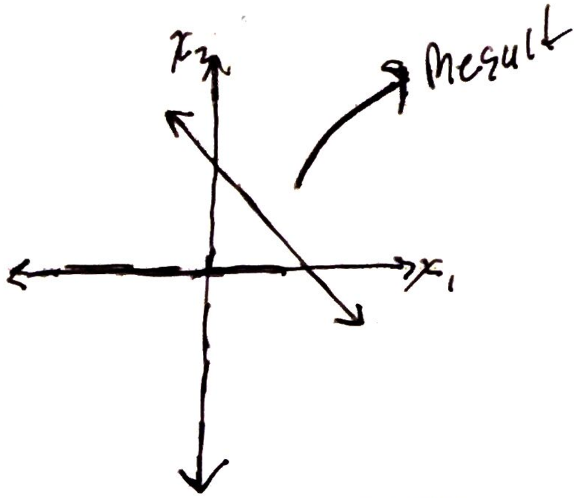
④ Repeat 2 and 3 for $i = 1, \ldots, n$

⑤ Repeat 2 to 4 until no error or m iterations

The perceptron algorithm ($\mathcal{A}$) is proven to converge if the $\mathbb{D}$ is linearly separable, but may fail if $\mathbb{D}$ is not.

Perceptron Example

i	x_1	x_2	y
1	-1	-1	0
2	1	1	1



Process

$\vec{w}^{\,t=0} = \vec{0}$

$j=1$

$t=1$

$\hat{y}_1 = 1$, $e_1 = y_1 - \hat{y}_1 = 0 - 1 = -1$
- $\vec{w}_0^{\,t=1} = 0 + (-1)(1) = -1$
- $\vec{w}_1^{\,t=1} = 0 + (-1)(-1) = 1$
- $\vec{w}_2^{\,t=1} = 0 + (-1)(-1) = 1$

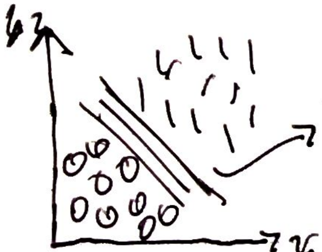
$\hat{y}_2 = 1$, $e_2 = y_2 - \hat{y}_2 = 0$
- $\vec{w}_0^{\,t=1} = -1 + (0)(1) = -1$
- $\vec{w}_1^{\,t=1} = 1 + (0)(1) = 1$
- $\vec{w}_2^{\,t=1} = 1 + (0)(1) = 1$

$j=2$

$\Rightarrow \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \geq 0 \Rightarrow -1 + x_1 + x_2 \geq 0 \Rightarrow x_2 \geq 1 - x_1$ Result above

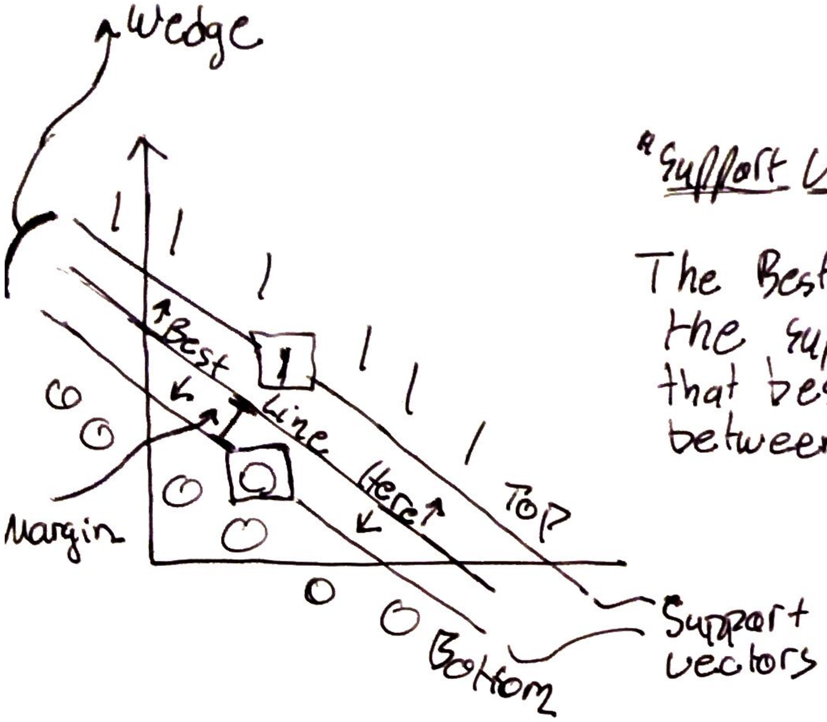
---

Likeart Scaling => Changing from actual metric to a scale, for example $1 \to 5$



These lines (1) of them are found by the perceptron

Somewhere in here is the best line.



"Support Vector Machine" Derivation

The Best Line lies between the support Vectors. The line that best discriminates must be between our support vectors

$$\mathcal{H} = \{ \mathbb{1}_{\vec{w}\cdot\vec{x}-b\geq 0} : \vec{w} \in \mathbb{R}^p, b \in \mathbb{R}$$

$\vec{w}\cdot\vec{x} - b = 0 \Rightarrow$ Hesse Normal Form of a line

$\ell_1 \Rightarrow x_2 = 2x_1 + 3 \Rightarrow \underbrace{\begin{bmatrix} 2 \\ -1 \end{bmatrix}}_{\vec{w}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \underbrace{-3}_{b} = 0$

↳ $2x_1 - x_2 + 3 = 0$

$\vec{w}$ is $\perp$ to ℓ and is called the <u>normal vector</u>

Let $\vec{w}_0 = \dfrac{\vec{w}}{\|\vec{w}\|} \Rightarrow$ unit vector in direction of $\vec{w}$

Let $\vec{z}$ be the vector on ℓ in the $\vec{w}_0$ direction

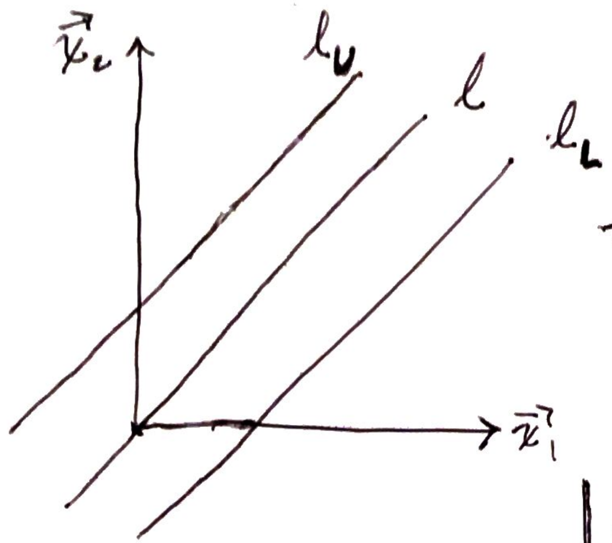
↳ This means $\vec{z} = \alpha\vec{w}_0 \mid \exists \alpha \in \mathbb{R}$, and b/c $\vec{z} \in \ell$, then

$\rightarrow \vec{w}\cdot\vec{z} - b = 0$

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

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

$$\Rightarrow \alpha = \frac{b}{\|\vec{w}\|} \Rightarrow \boxed{\vec{z} = \frac{b}{\|\vec{w}\|}\,\vec{w}_0}$$



(Lookup Diagram for Hesse Normal Form of a Line)

$$l_U : \vec{w}\cdot\vec{x} - (b+\delta) = 0$$

$$l_L : \vec{w}\cdot\vec{x} - (b-\delta) = 0$$

$$\Rightarrow \vec{z}_U = \frac{b-\delta}{\|\vec{w}\|}\cdot\vec{w}_0$$

$$\Rightarrow \vec{z}_L = \frac{b+\delta}{\|\vec{w}\|}\cdot\vec{w}_0$$

Let m be width of wedge

$$m = \|\vec{z}_U - \vec{z}_L\|$$

$$= \left\|\frac{b+\delta}{\|\vec{w}\|}\vec{w}_0 - \frac{b-\delta}{\|\vec{w}\|}\vec{w}_0\right\|$$

$$= \left\|\frac{1}{\|\vec{w}\|}\left((b+\delta)-(b-\delta)\right)\vec{w}_0\right\|$$

$$= \left\|\frac{1}{\|\vec{w}\|}2\delta\vec{w}_0\right\| = \frac{2\delta}{\|\vec{w}\|}\,\|\vec{w}\| = \frac{2\delta}{\|\vec{w}\|}$$