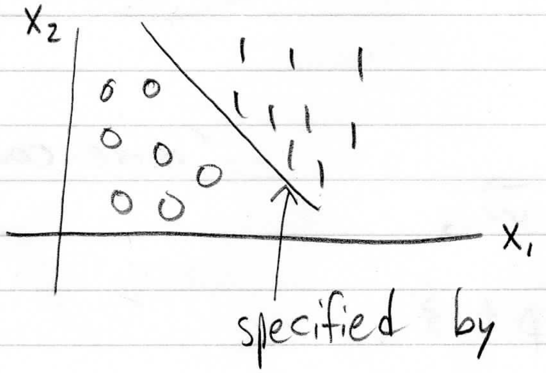
2/6

$Y = \{0, 1\}$

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

"Line of Linear Discrimination"

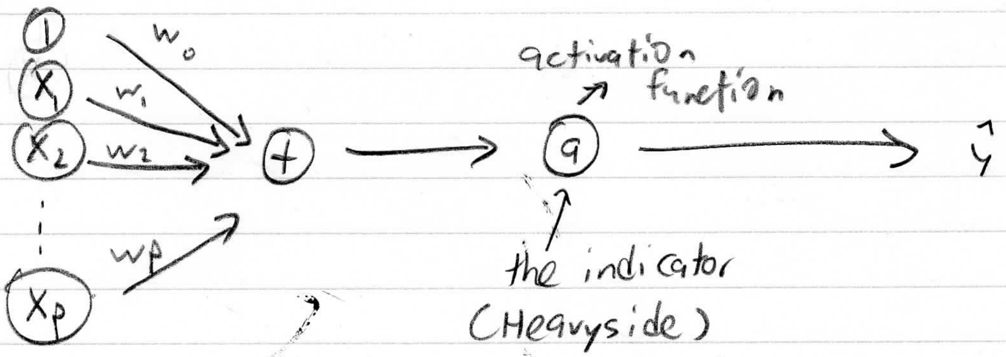
for $p = 2$,



"Linear seperability"

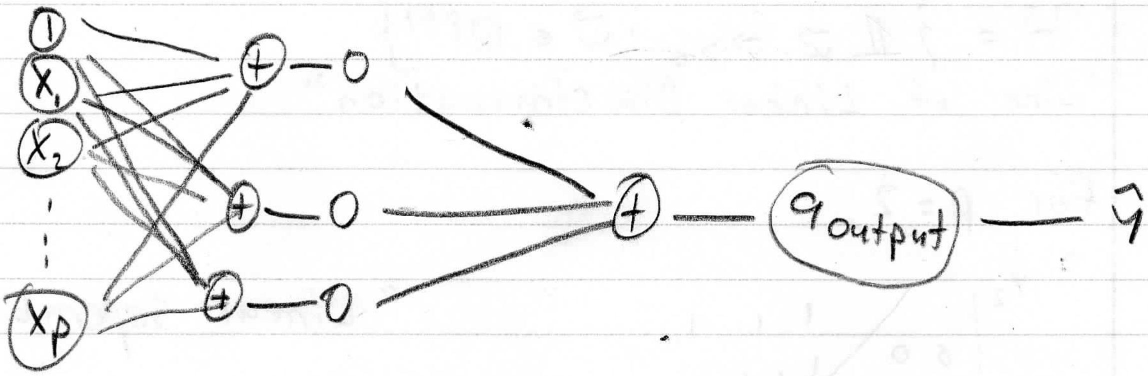
If data is linearly seperable, the perceptron algorithm is guaranteed to converge.

Input layer | Output layer | Prediction



Layers

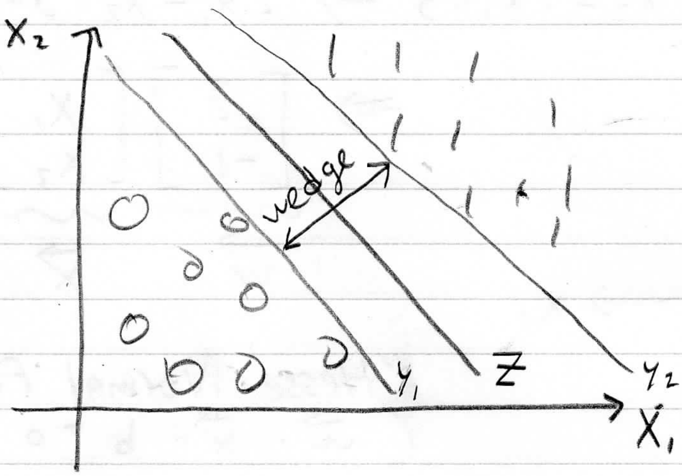
Input Hidden Output



'more complicated'

$\vec{w}_{11}, \vec{w}_{21}, \vec{w}_{31}, \vec{w}_{12}$

\# of parameters : $3p + 3$



$$X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix}$$

Z is what we want (average of the two lines that seperate the 0's + 1's).

y_1, y_2 are support vector.

- "Maximum Margin Hyperplane"

Proven in 1998 to be the optimal linear classifier

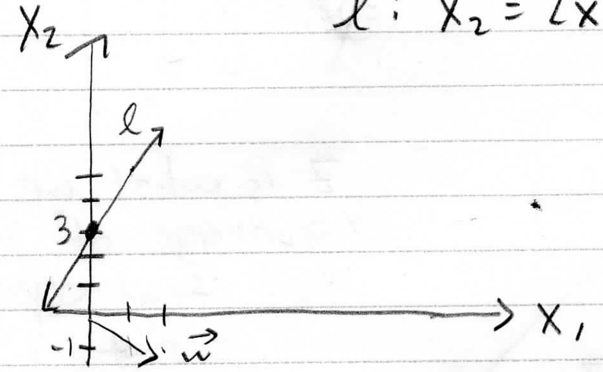
A: support vector machine (SVM)

$$\vec{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_p \end{bmatrix}, \quad w_0 = b,$$

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

↑ p measurements

$$\ell : x_2 = 2x_1 + 3 \Rightarrow 2x_1 - x_2 + 3 = 0$$



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

"Hesse Normal Form"

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

$\vec{w}$ is the "normal" vector is $\perp$ to ℓ