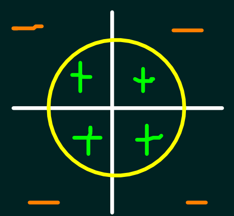
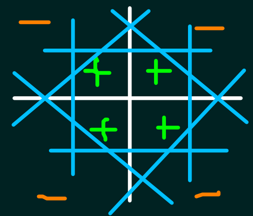
SUMMARY

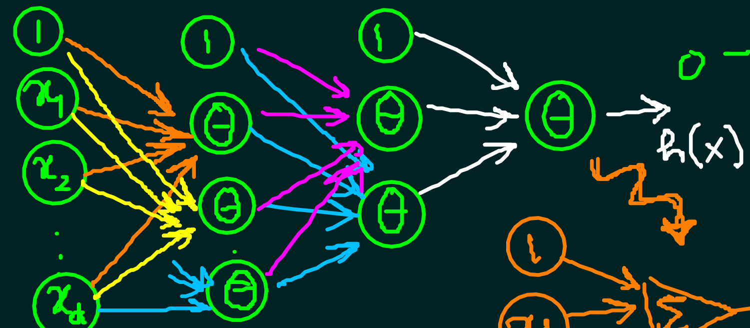
▷ Multi-layer Perceptrons:

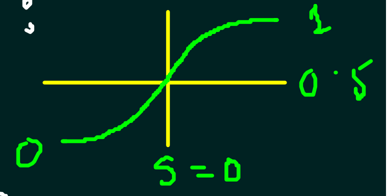




logical combination of Perceptrons

▷ Neural Network:





$$h(x) = x_1^{(L)}$$

$$x_j^{(l)} = \theta\left(\sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)}\, x_i^{(l-1)}\right)$$

Where, $\theta(s) = \frac{1}{1+e^{-s}}$

$$\theta'(s) = \theta(s)\left[1-\theta(s)\right]$$

▷ Backpropagation Procedure:

$$\frac{\partial E(w)}{\partial w_{ij}^{(l)}} = \underbrace{\cancel{\frac{\partial E(w)}{\partial s_j^{(l)}}}}_{\boxed{\delta_j^{(l)}}} \times \underbrace{\cancel{\frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}}}}_{x_i^{(L-1)}}$$

Compute Recursively

$$\boxed{\begin{array}{c} 1 \le l \le L \\ 0 \le i \le d^{(l-1)} \\ 1 \le j \le d^{(l)} \end{array}}$$

↳ Base Case (last layer): $\delta_1^{(L)} = \frac{\partial E(w)}{\partial s_1^{(L)}} = -\left(y_n - x_1^{(L)}\right) \cdot x_1^{(L)} \cdot \left(1 - x_1^{(L)}\right)$

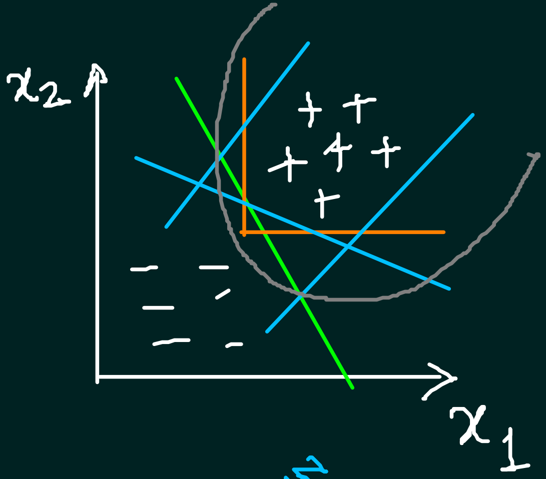
[as, $E(w) = \frac{1}{2}\left(y_n - x_1^{(L)}\right)^2$ and $\theta(s) = x_1^{(L)}$]

↳ Recursive (Backwards):

$$\boxed{\delta_i^{(l-1)}} = \frac{\partial E(w)}{\partial s_j^{(l-1)}} = \sum_{j=1}^{d^{(l)}} \underbrace{\cancel{\frac{\partial E(w)}{\partial s_j^{(l)}}}}_{\boxed{\delta_j^{(l)}}} \times \underbrace{\cancel{\frac{\partial s_j^{(l)}}{\partial x_i^{(l-1)}}}}_{\boxed{w_{ij}^{(l)}}} \times \underbrace{\cancel{\frac{\partial x_i^{(l-1)}}{\partial s_j^{(l-1)}}}}_{\boxed{\theta'\left(s_i^{(l-1)}\right)}}$$

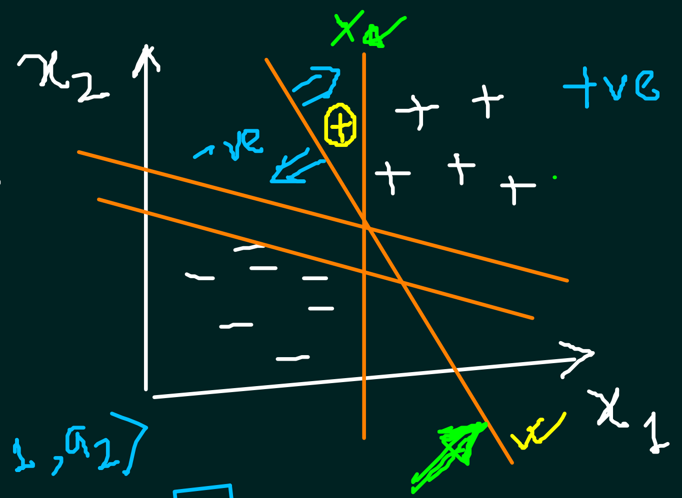
$$\Longrightarrow \delta_i^{(l-1)} = x_i^{(l-1)}\left[1 - x_i^{(l-1)}\right] \times \sum_{j=1}^{d^{(l)}} w_{ij}^{(l)} \times \delta_j^{(l)}$$

Hence, $\Delta w_{ij}^{(l)} = -\eta \cdot x_i^{(l-1)} \cdot \delta_j^{(l)}$ where: $\delta_i^{(l)} = x_i^{(l-1)}\left[1 - x_i^{(l-1)}\right] \sum_{j=1}^{d^{(l)}} w_{ij}^{(l)}\, \delta_j^{(l)}$



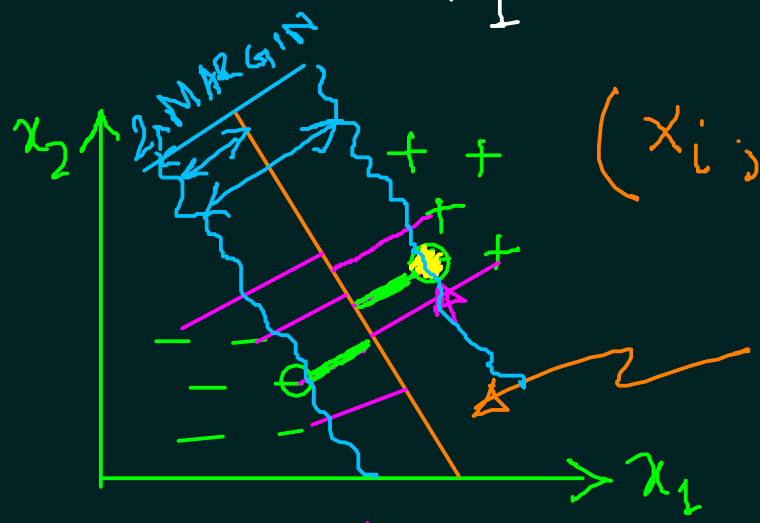
$$f: X \to Y = \begin{cases} +1 \\ -1 \end{cases}$$

Discriminant Analysis
(linear)



$$x^{new} = \langle a_1, a_2 \rangle$$

$$\begin{cases} w_1 a_1 + w_2 a_2 + b \geq 0 \to \boxed{+1} \\ \qquad < 0 \to \boxed{-1} \end{cases} \qquad w_1 x_1 + w_2 x_2 + b = 0$$



$$\left(x_i, \overset{y_i}{\pm}\right)$$

$$w^T x + b = 0 \qquad (w_1 x_1 + w_2 x_2 + b = 0)$$

$$\geq 0 \to +ve, \quad < 0 \to -ve$$

$$d_i = \frac{|w_1 x_{i1} + w_2 x_{i2} + b|}{\sqrt{w_1^2 + w_2^2}} \quad \therefore \quad \boxed{y_i (w_1 x_{i1} + w_2 x_{i2} + b) \geq \cancel{0}\ 1}$$

$$\geq 1$$

$$\max_{w_1, w_2, b} \left[\min_i (d_i)\right] = \max \left[\min_i \frac{\overset{1}{\cancel{|w_1 x_{i1} + w_2 x_{i2} + b|}}}{\sqrt{w_1^2 + w_2^2}} \right]$$

$$\boxed{\max \left(\frac{1}{\sqrt{||w||}} \right)} \to ||w|| = w^T w$$

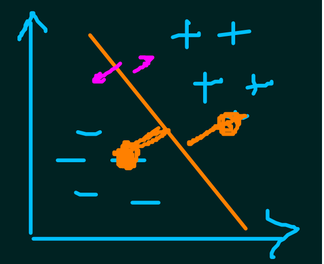
$$\begin{cases} 2x + 4y + 12 = 0 \\ 4x + 8y + 24 = 0 \\ x + 2y + 6 = 0 \end{cases}$$

PRIMAL OPT. PROBLEM:

Minimize $\frac{1}{\sqrt{\|w\|}} \Rightarrow \text{Max}\left(\frac{1}{2} w^T w\right) = L$

Eg.

Subject to. $y_i(w^T x_i + b) \geq 1 -$



DUAL OPT. PROBLEM: $x_i \leftrightarrow \alpha_i$ (lagrange multiplier) [Support Vector Machine]

KKT

Constr.: $\boxed{\alpha_i \geq 0}$

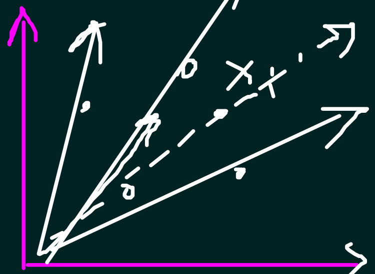
Maximize, $L = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i \left(y_i(w^T x_i + b) - 1\right)$

$\frac{\partial L}{\partial w} = 0 \rightarrow w - \sum_{i=1}^{N} y_i \alpha_i x_i = 0 \Rightarrow$ A $\boxed{w = \sum_{i=1}^{N} y_i \alpha_i x_i}$

$\frac{\partial L}{\partial b} = 0$ B $\boxed{\sum_{i=1}^{N} \alpha_i y_i = 0}$

$w^T x + b = 1$

$b = 1 - w^T x$



$w \leftarrow$ weight $\quad b \leftarrow$ bias

A + B

$L = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i y_i \alpha_j y_j (x_i \cdot x_j)$

$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j \underbrace{y_i y_j (x_i \cdot x_j)} - b \cancel{\sum_{i=1}^{N} \alpha_i y_i}^{\,0} + \sum_{i=1}^{N} \alpha_i$

Therefore, $W = \sum_{i=1}^{N} \alpha_i y_i X_i$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$

$$L = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \underbrace{y_i y_j (X_i \cdot X_j)}_{\mathcal{H}}$$

$\mathcal{H}$ (Hessian Matrix)

$$= \Lambda U^T - \frac{1}{2} \Lambda \mathcal{H} \Lambda^T$$

$\Lambda = [\alpha_1 \; \alpha_2 \cdots \alpha_N]_{1 \times N}$, $U = [1 \; 1 \cdots 1]_{1 \times N}$

$$\mathcal{H} = \begin{bmatrix} \cdots & \cdots \\ & y_i y_j x_i \cdot x_j \end{bmatrix}$$

Solved using Quadratic Programming Tech (w.r.t. α)

After solving this:

$\boxed{\alpha_i \text{ values}} \geq 0$

$W = \sum_{i=1}^{N} \alpha_i y_i X_i$ ← eff.

and $b = 1 - W^T x$

SUPPORT LINE (SVM)