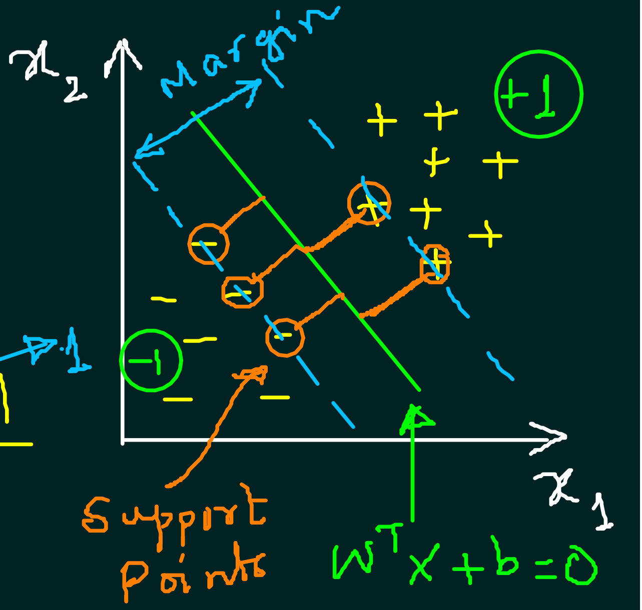
Support Vector Machine (SVM): SUMMARY



↳ linear discriminator, $\boxed{W^T x + b = 0}$

$\therefore \; \forall i, \langle x_i, y_i \rangle$ from TE, $y_i(W^T x_i + b) \geq 0$

↳ Margin of Line $\coloneqq \min\limits_i d_i = \min\limits_i \dfrac{|W^T x_i + b|}{\|W\|}$ → 1

▷ Primal Optimization Problem:

Maximize $\left[\dfrac{1}{\|W\|^2}\right]$ i.e. Minimize $L = \frac{1}{2} W^T W$

subject to, $y_i(W^T x_i + b) \geq 1$

▷ Dual Optimization Problem: ($\alpha_i \leftarrow$ Lagrange Multipliers)

Minimize $L = \frac{1}{2} W^T W - \sum_{i=1}^{N} \alpha_i \left(y_i (W^T x_i + b) - 1 \right)$ such that, $\alpha_i \geq 0$

$\dfrac{\partial L}{\partial W} = 0 \Rightarrow W = \sum_{i=1}^{N} \alpha_i y_i x_i$, and $\dfrac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$

$\therefore L_{min} = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i \cdot x_j = \Lambda U^T - \frac{1}{2} \Lambda \mathcal{H} \Lambda^T$

$[\alpha_1 \ldots \alpha_N]$ $[1 \ldots 1]$ Hessian Mat $[y_i y_j x_i x_j]$

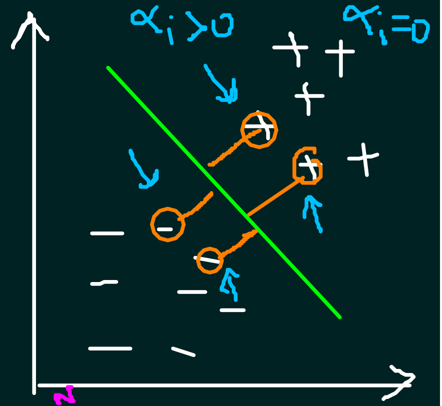
Hence, $W = \sum_{i=1}^{N} \alpha_i y_i x_i$ and $b = 1 - W^T x_i$ (Solved using Q.P.)

▷ KKT Theorem:

Primal Opt Problem → **Dual Optimization Problem**

$$\text{Min } \frac{1}{2} w^T w \quad \text{s.t } y_i (w^T x_i + b) \geq 1 \quad \forall (x_i, y_i)$$

Const | Opt.



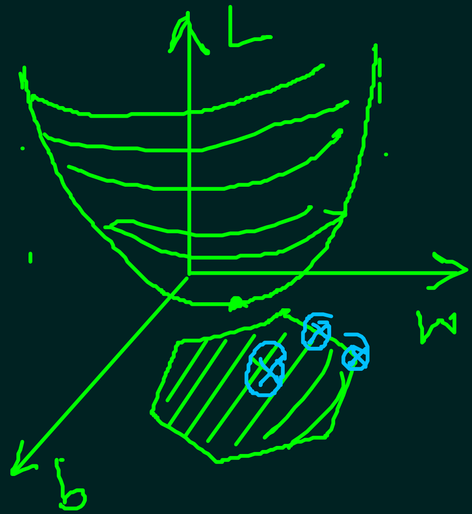
$y_i (w^T x_i + b) > 1$ ⟷ $\alpha_i = 0$

(and) $y_i (w^T x_i + b) = 1$ ⟷ $\alpha_i > 0$

boundary point

interior point

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

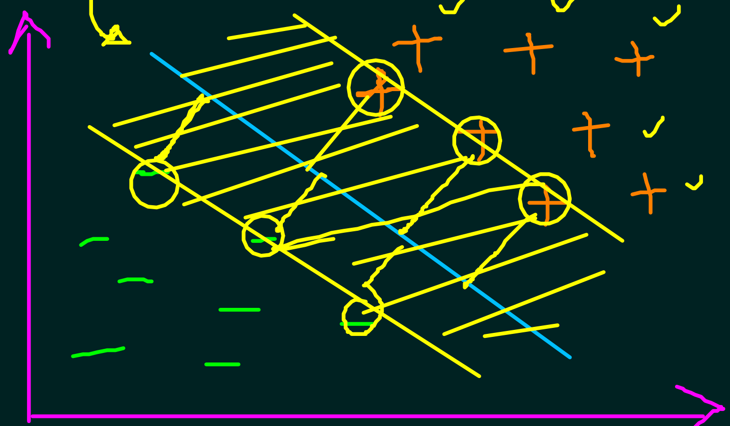


$$\boxed{w = \sum_{i \in SV} \alpha_i y_i x_i}$$ weigh

efficient in Comp. ⇐

$$W = \sum_{i=1}^{N} \alpha_i y_i x_i$$

only counted for those $\alpha_i > 0$



$$\boxed{b = 1 - w^T x_i}$$

bias

$\frac{x}{3} + \frac{1}{2} y + \frac{2}{3} = 0$

$2x + 3y + 4 = 0$

$x + \frac{3}{2} y + 2 = 0$

$x_1 = 100$ $\quad x_2 = 100$

$$\boxed{|\vec{w}_1 x_1 + \vec{w}_2 x_2 + \vec{b}|}$$

▷ Steps of SVM :

[·] [ ]

↳ Find Hessian Matrix $\mathcal{H}_{ij} = \{ y_i y_j X_i^T \cdot X_j \}$

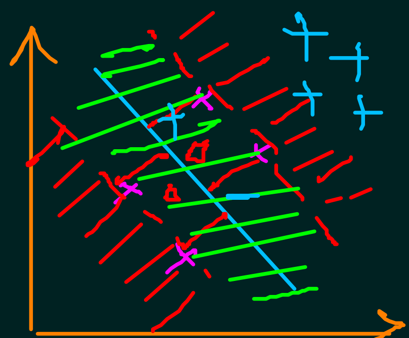
↳ Solve Min, $L = \Lambda U^T - \frac{1}{2} \Lambda \mathcal{H} \Lambda^T$

s.t. $\alpha_i \geq 0$

Train

↳ Find $w = \sum_{i \in SV} \alpha_i y_i X_i$ ✓

and $b = 1 - W^T X_i$ ✓



Classify

→ $X^{new} = \langle x_1^n, x_2^n, \dots, x_k^n \rangle$

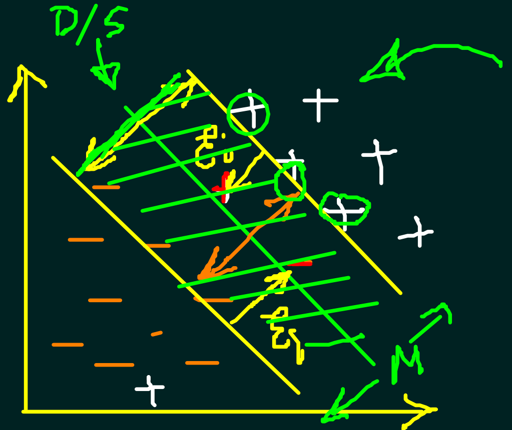
↳ check for $\text{sign}(W^T X^{new} + b) \gtrless 0$

$= \text{sign}\left(\sum_{X_i \in SV} y_i (X_i \cdot X_{new}) \right)$

▷ Issues :

↳ Noisy / Overlap Classes

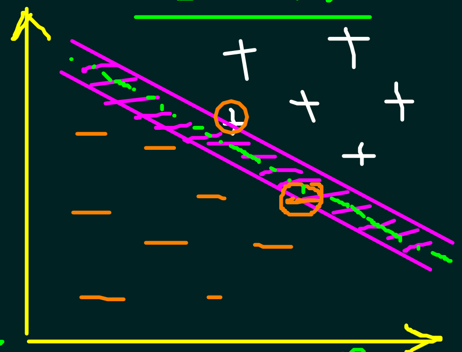
↳ Non-linear → Multi-class



Soft SVM Margin:

$(M)\ \frac{M}{2} \geq \varepsilon_i \geq 0 \leftarrow$ slack

HARD Margin



$\varepsilon_i = 0$

Primal Opt:

$$\text{Min } L = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \varepsilon_i^2$$

$$\text{s.t. } \quad y_i (w^T x_i + b) \geq 1 - \varepsilon_i$$

KKT

Dual Opt. Prob $\left\{ C \geq \alpha_i \geq 0 \quad \mu_i \geq 0 \right.$

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

$\boxed{\text{Q.P.}}$

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

$$\Rightarrow w = \sum_i \alpha_i y_i x_t$$

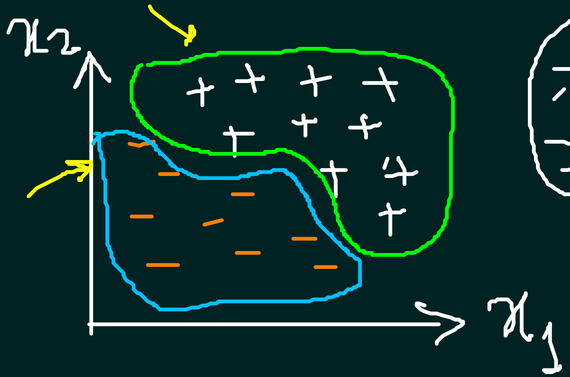
$$\frac{\partial L}{\partial b} = \sum_i \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \varepsilon_i} = C - \alpha_i - \mu_i = 0$$

$$\Rightarrow C = \alpha_i + \mu_i$$

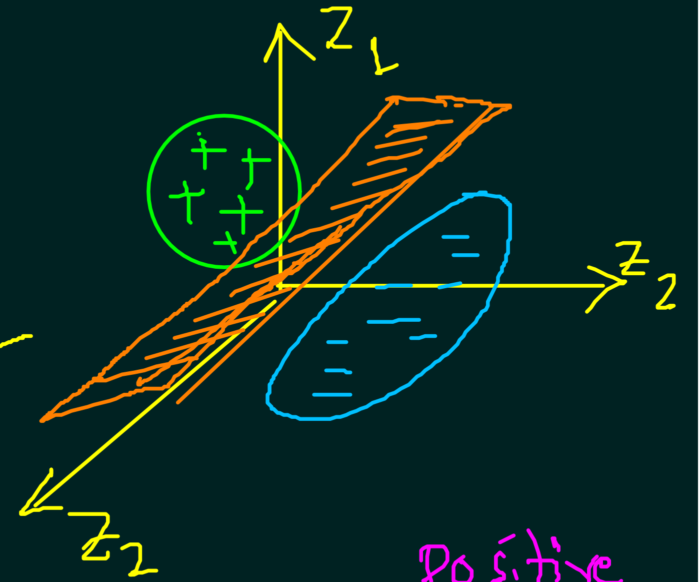
$C \leftarrow$ generalization const $\Big\}$ Train vs Test error (margin) controlled using C

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



Transform

$$Z \leftarrow \phi(X)$$



Non-linear (discrimination)

$$\boxed{L = \underline{\Lambda} U^T - \frac{1}{2} \underline{\Lambda} \boxed{H} \underline{\Lambda}^T}$$

$\alpha_i \rightarrow \langle x_i, y_i \rangle$

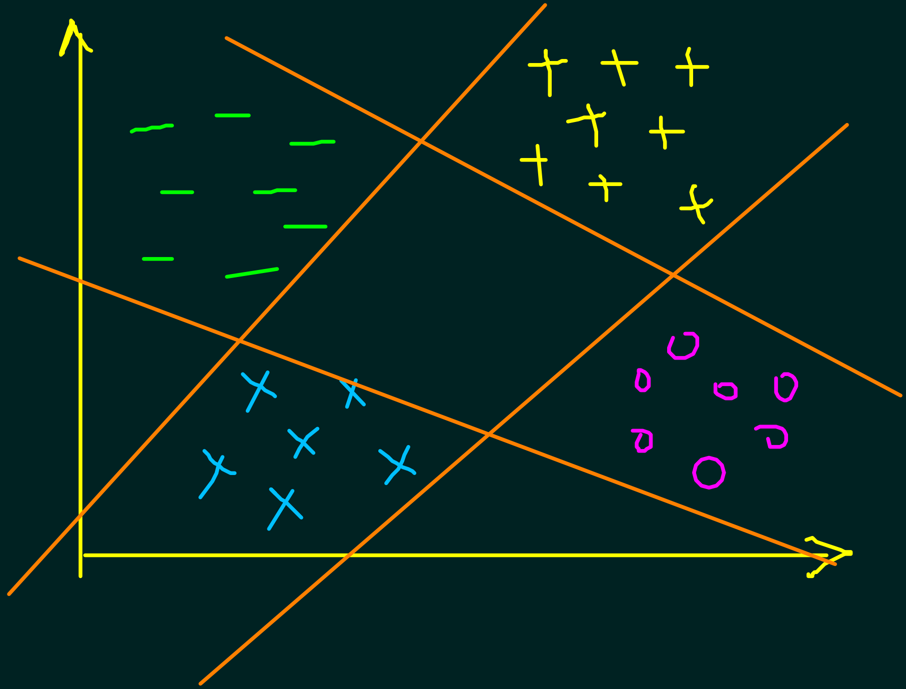
Positive definite

$\Rightarrow K(x_i, x_j) \leftarrow$ Kernel function (SVM)

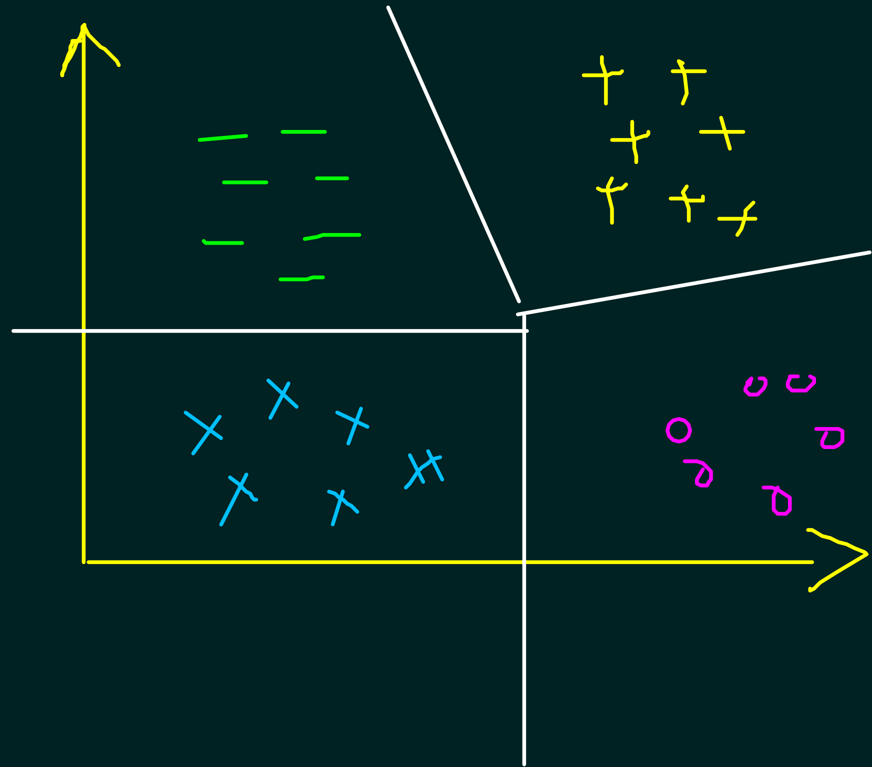
$$H = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ & & y_i y_j \boxed{\cancel{x_i . x_j}} & & \\ & & K(x_i . x_j) & & \end{bmatrix}$$

- $\boxed{\text{Linear Kernel} = X_i . X_j}$
- Quadratic Kernel = $(1 + X_i . X_j)^Q$
- exp. Kernel = $e^{(\dots)}$ or ---- ...
  - Gaussian Nature $\Rightarrow$ **Radial Basis Function**

▷ Multi class Classification:



- One-against-All
- One-to-one



SUMMARY

- SV — Soft / Hard
- D.L — Soft / Hard
- Margin
- Kernel Function
- Multiclass