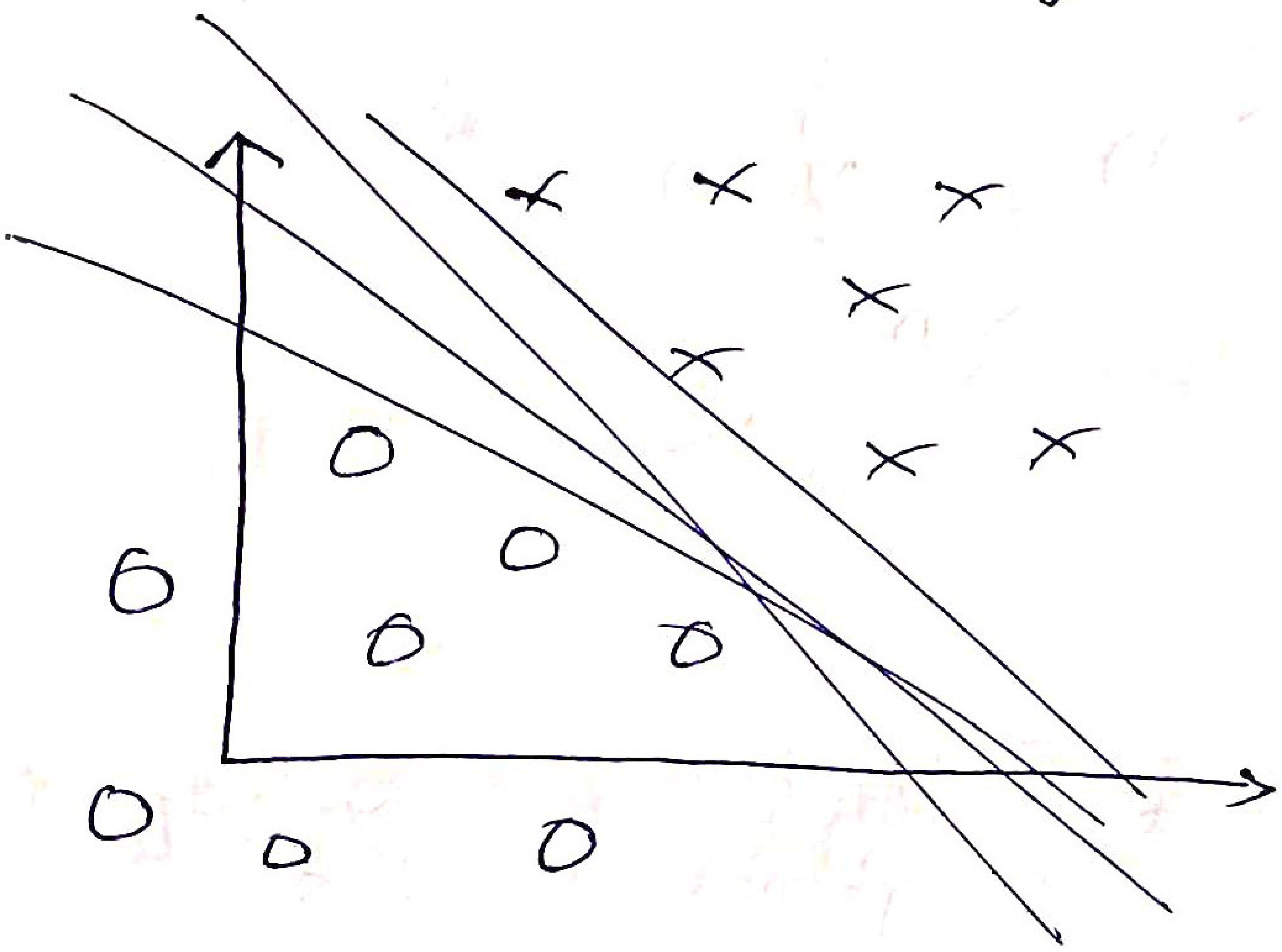
SVM

(Support Vector Machine)



⇒ It is an algorithm for performing **binary classification**.

↳ We take feature vector X map it to high dimension vector.

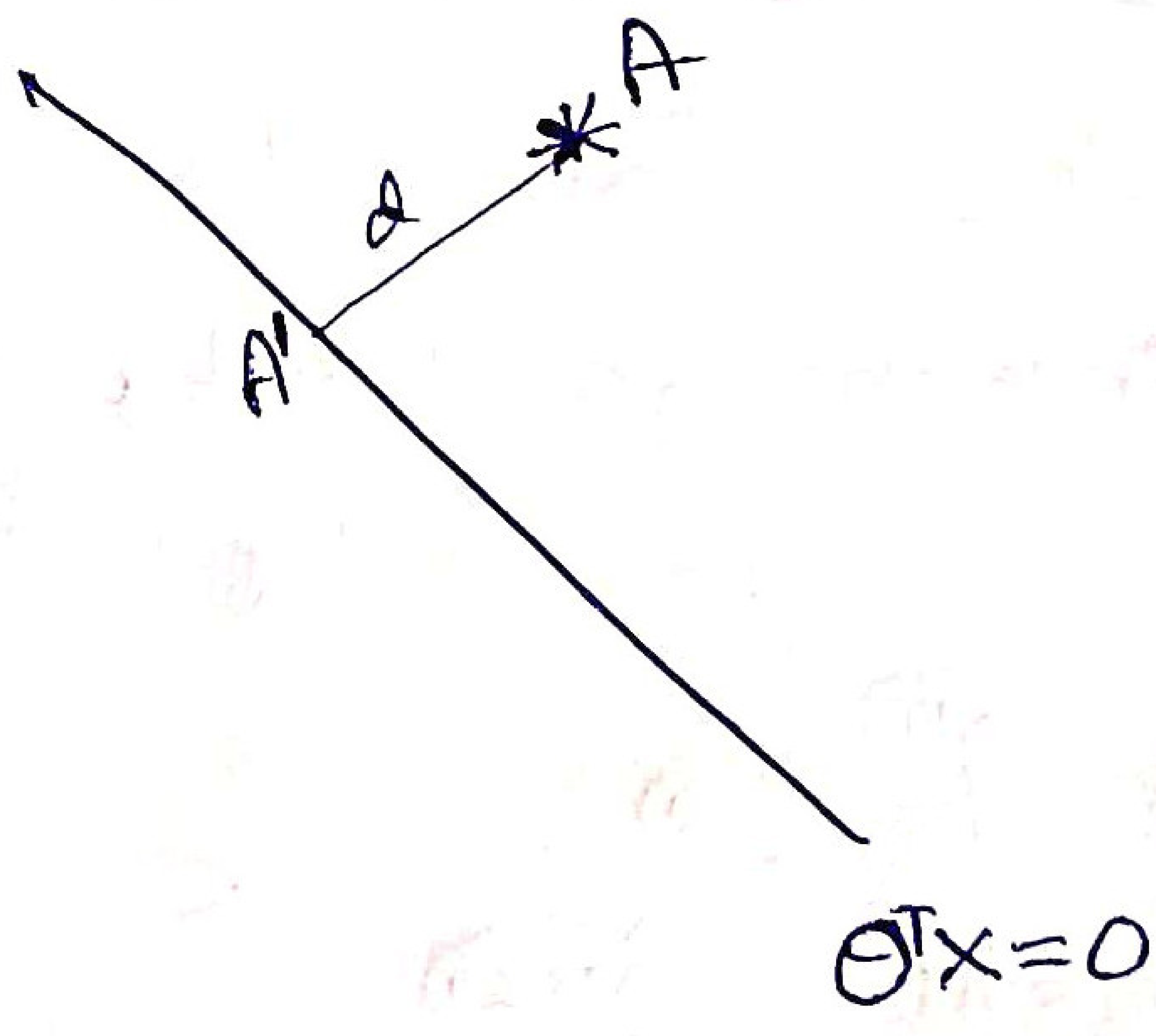
$$X \longrightarrow X^*$$

Optimal Margin Classifier

⇒ Assumption: Inputs can be seperated into different classes by a linear decision boundary.

margin

- → Functional Margin $\theta^T A$
- → Geometric Margin



$\Rightarrow$ Let $\{(x^{(i)}, y^{(i)}) \mid x^{(i)} \in R^n, y^{(i)} \in \{-1, 1\}, i \in \{1, \dots m\}\}$ be the training set.

Objective: Find linear decision boundary ~~class~~ separating +ve & and -ve examples.

Assumption: Inputs are Linearly Seperable.

$\Rightarrow$ Let the seperating hyperplane be parametrized as

$$w^T x + b = 0, \quad w \in R^n \;\&\; b \in R$$

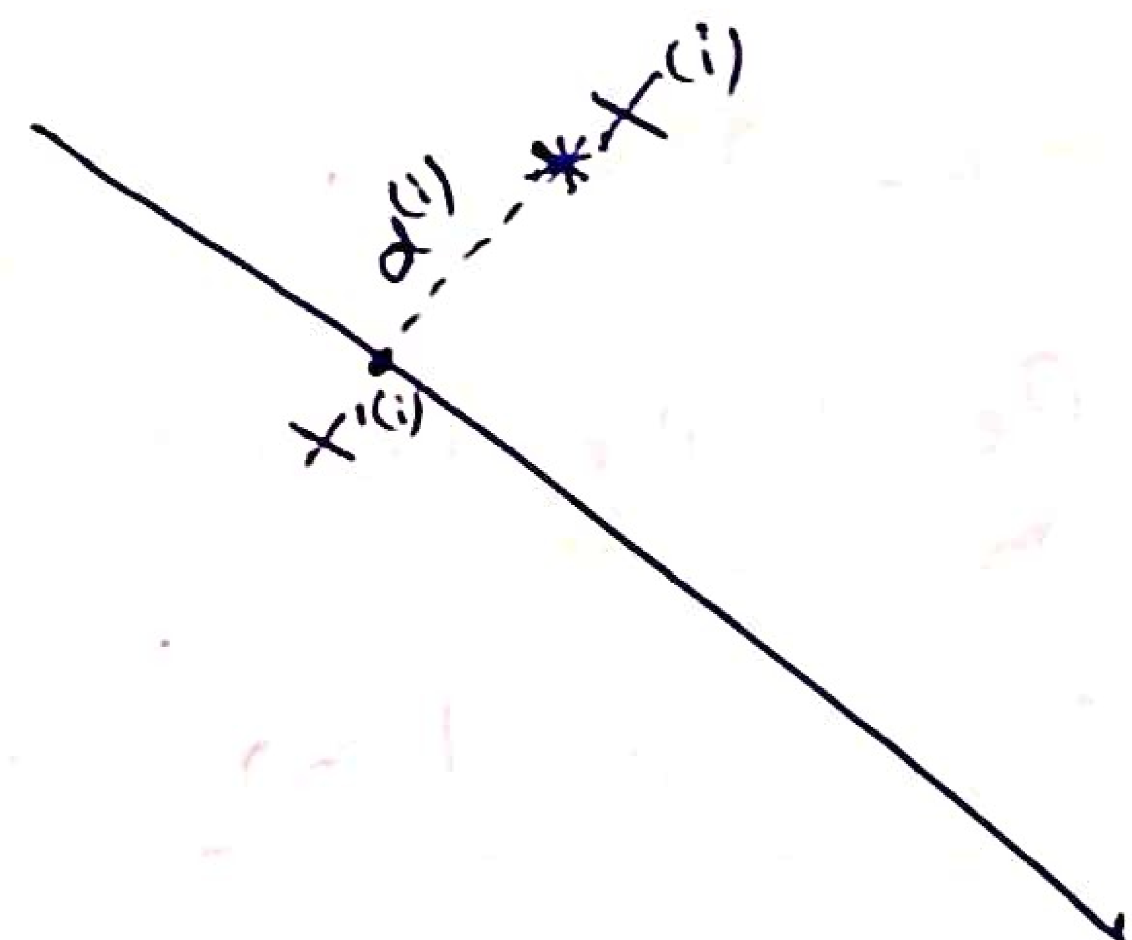
$\Rightarrow$ Let us define Margin for a data point $(x^{(i)}, y^{(i)})$:

① $\rightarrow$ Functional Margin

$y^{(i)} \theta^T x^{(i)}$

or $\qquad \theta = \begin{bmatrix} b \\ w \end{bmatrix}$

$$\boxed{y^{(i)} [w^T x^{(i)} + b]}$$



② $\rightarrow$ Geometric Margin

Margin

$\rightarrow$ Gives a sense of how far a point is from separating line. (decision boundary) (No direction)

$$x'^{(i)} = x^{(i)} - d^{(i)} y^{(i)} \frac{w}{\|w\|_2}$$

$$w^T x'^{(i)} + b = 0$$

$$w^T x^{(i)} - d^{(i)} y^{(i)} \frac{w^T w}{\|w\|_2} + b = 0$$

$\|w\|_2^2$

$\|w\|_2$

$$d^{(i)} y^{(i)} \|\omega\|_2 = \omega^T x^{(i)} + b$$

$$\Rightarrow \boxed{d^{(i)} = y^{(i)} \left(\left[\frac{\omega}{\|\omega\|_2} \right]^T x^{(i)} + \frac{b}{\|\omega\|_2} \right)}$$

(Euclidian)

⇒ Geometric margin is the absolute distance of a data point from the decision boundary.

★ Geometric Margin of training Set

$$\min_i d^{(i)}$$

or

$$\boxed{\min_i y^{(i)} \left(\left[\frac{\omega}{\|\omega\|_2} \right]^T x^{(i)} + \frac{b}{\|\omega\|_2} \right)}$$

⇒ Optimal margin Classifier, finds the decision boundary that maximizes the geometric margin.

⇒ The above can be formulated as the following optimization Problem :

$$\cancel{\max_{\omega, b} \min_i \frac{y^{(i)} (\omega^T x^{(i)} + b)}{\|\omega\|_2}} \quad \forall i = \{1, \dots m\}$$

$$\boxed{\begin{aligned} &\max_{\omega, b, \gamma} \gamma \\ &\text{st} \quad \frac{y^{(i)} (\omega^T x^{(i)} + b)}{\|\omega\|_2} \geq \gamma \quad \forall i = \{1, \dots m\} \end{aligned}}$$

⇒ In this optimization problem, we can scale $\|w\|_2$ whatever we want without changing the decision boundary.

⇒ So, let $\|w\|_2 = 1/\gamma$

⇒ $$\max_{w,b} \frac{1}{\|w\|_2}$$

$$\text{st } y^{(i)}(w^T x^{(i)} + b) \geq 1 \quad \forall i = 1, \cdots m$$

$$\boxed{\begin{aligned} &\min_{w,b} \frac{1}{2}\|w\|_2^2 \\ &\text{st. } y^{(i)}(w^T x^{(i)} + b) \geq 1 \quad \forall i = 1, \cdots m \end{aligned}}$$

{ This can be easily solved using QP }

★ <u>Expressing Optimal Margin classifier optimization in terms of $\langle x^{(i)}, x^{(j)} \rangle$</u>

⇒ Assumption

$$w = \sum_{i=1}^{m} \alpha_i x^{(i)}$$

{ Representer Theorem justifies this assumption }

$$\boxed{w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}}$$

{ To make the maths, little bit easier }

⇒ When we plug this ω to the OMC problem we get:

$$\boxed{\begin{aligned}&\min_{\alpha, b} \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y^{(i)} y^{(j)} \langle x^{(i)}, x^{(j)} \rangle \\ &\text{St. } y^{(i)}\Big(\sum_j \alpha_j y^{(j)} \langle x^{(i)}, x^{(j)} \rangle + b\Big) \geq 1\end{aligned}}$$

⇒ So entire optimization problem can be written in terms of $\langle x^{(i)} x^{(j)} \rangle$.

⇒ At the prediction stage also result can be expressed in terms of $\langle x^{(i)}, x^{(j)} \rangle$.

$$h_{\omega,b}(x) = g(\omega^T x + b)$$
$$= g\Big(\sum_i \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b\Big)$$

★ <u>Kernel Trick</u>

⇒ Applying kernel trick to optimal margin classifier is referred to as SVM.

⇒ Applying kernel trick:

1. Write algorithm in terms of $\langle x^{(i)}, x^{(j)} \rangle$
2. Let there be some mapping from $x \rightarrow \phi(x)$
3. Find a way to compute $K(x, z) = \phi(x)^T \phi(z)$
4. Replace $\langle x, z \rangle$ in algorithm with $K(x, z)$