U19CS009

Brijesh Rohit

Q-6 D) In SVM, which is a linear classifier we define a linear classifier as

$$h(x) = \text{sign}(w^T x + b).$$

and the binary classification settings are labelled +1 & -1

→ If it is possible to divide a data set, multiple hyperplane $h(x)$ are possible, but the one that maximises the distance of closest data points from both class is best.

Say we have point x from hyperplane ~~H = {x | wᵀx~~

$$H = \{x \mid w^T x + b = 0\}.$$

$$x^P = x - d$$

d → parallel ~~distance~~ to w and is a vector

$\therefore d = \alpha w$ for some $\alpha \in \mathbb{R}$

$\because x^P \in H$. x^P → is projection of x on H.

$$\therefore w^T x^P + b = w^T(x-d) + b = w^T(x - \alpha w) + b = 0$$

$$\therefore \alpha = \frac{w^T x + b}{w^T w}.$$

$$\therefore d = \sqrt{d^T d}$$

$$= \sqrt{\alpha^2 w^T w} = \frac{|w^T x + b|}{\sqrt{w^T w}}$$

$\therefore$ margin of H with respect to D: $\gamma(w, b)$

$$= \min_{x \in D} \left| \frac{w^T x + b}{\sqrt{w^T w}} \right|.$$

U19CS009

Brijesh Rohit

Now for max. $\gamma(w,b)$.

$$\rightarrow \max_{w,b} \frac{1}{\cancel{\|w\|}} \quad \frac{1}{\sqrt{w^T w}} \min_{x_i \in D} |w^T x_i + b|$$

we can add this rescalling as an equality constraint.

$$\therefore \max_{w,b} \frac{1}{(\sqrt{w^T w})^2} \cdot 1 = \min_{w,b} (\sqrt{w^T w})^2.$$

$$= \min_{w,b} w^T w$$

-> If w is weight vector, than a functional margin of 1 implies

$$\langle w \cdot x^+ \rangle + b = +1 \quad (w \cdot x^-) + b = -1$$

$$\therefore \gamma = \frac{1}{\|w\|_2}$$

for linearly sepreable ~~tra~~ sample

$$\text{minimize}_{w,b} \quad \langle w \cdot w \rangle$$