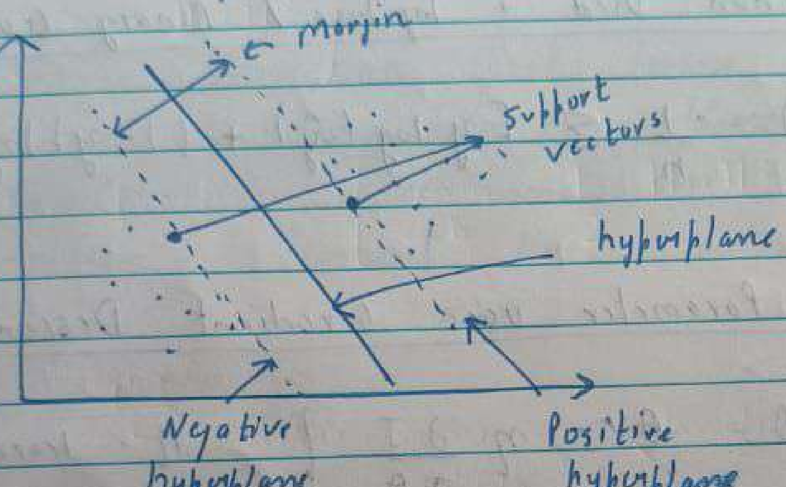
Date: __ . __ . __

SVM (Support vector Machine)

It is a supervised ML algo, used for both classification & regression tasks. It tries to find best boundary known as hyperplane that separate different classes in data.



Support vectors → These are the closest data points to hyperplane that define class boundary.

Hyperplane → It is a plane that separate different classes.

Main Objective → Goal is to maximize the margin b/w two classes. The larger margin better model performance.

Kernel → A function that maps data to a higher dimensional space enabling SVM to handle non-linear separable data.

Kernel → Linear, Polynomial, RGP.

Standard Distance. plane $ax + by + cz + d = 0$

point (x_0, y_0, z_0).

$$d = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}$$

Our SVM Plane.

$$\boxed{wx + b = 0}$$

$$\cancel{w_1 x_1 + w_2 x_2 + \ldots w_n x_n + b = 0}$$

SVM Distance Formula $= \left\{ \frac{|w \cdot x_0 + b|}{||w||} \right\}$

SVM planes (+ve & -ve) $\rightarrow$ $wx + b = +1$

$wx + b = -1$

SVM Plane $\rightarrow$ $wx + b = 0$

in 2D $\rightarrow$ $w_1 x + w_2 y + b = 0$

Support vector $\rightarrow$ $w_1 x + w_2 y + b = 1$ (+ve one)

any point

(x_8, y_8) on line $\rightarrow$ $w_1 x_8 + w_2 y_8 + b = 1$

Finding Distance b/w Point & Plane (SVM) $\Rightarrow$ margin of one side

$$d = \frac{|w_1 x_8 + w_2 y_8 + b|}{\sqrt{w_1^2 + w_2^2}}$$

$$\left\{ d = \frac{1}{||w||} \right\}$$

As $\left\{ \sqrt{w_1^2 + w_2^2} = ||w|| \text{ vector form} \right\}$

Same distance from negative Plane side.

$$\boxed{\text{Total margin} = \frac{2}{||w||}}$$

$\rightarrow$ We have to maximize this.

apsara

Date: __.__.__

Showing how $\max \frac{2}{||w||}$ is same as $\min \frac{1}{2} ||w||^2$

$$\max \frac{2}{||w||} \overset{\text{Remove constant}}{\equiv} \max \frac{1}{||w||} \overset{\text{Reciprocal}}{\equiv} \max \frac{1}{||w||} \text{ same as } \min ||w||$$

$\min ||w||$ squaring not change min point.

$$\Downarrow$$

$\min ||w||^2$ Now, adding $\frac{1}{2}$ constant for math convinience

$$\Downarrow$$

$$\left\{ \frac{1}{2} \min ||w||^2 \right\}$$

☆ Objective function, $\left[\min_{w,b} \frac{1}{2} ||w||^2 \right]$ &

- Soft Margin SVM
  ↳ Allow some misclassification by introducing slack variable, soft margin SVM used when data is not perfectly seperable.

  In this we introduce hinge loss.

$$\left[\min_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \max \left(0, 1 - y_i (w.x_i + b)\right) \right]$$

- hard Margin SVM
  ↳ It is used when data is perfectly linearly separable.

$$\left[\min \frac{1}{2} ||w||^2 \right]$$

apsara

Date: __.__.____

As we are going to find best hyperplane $wx + b = 0$ learnable parameter are w & b (best values find task).

In this for optimization we use Lagrange eqn instead of Gradient descent as GD is not suitable for strictly constrained optimization.

Lagrang = objective - $\sum$ multiplier X Constraint.

$$\left\{ \frac{1}{2} ||w||^2 - \sum \alpha_i [y_i (w.x_i + b) - 1] \right\}$$

Key Points

→ Use svm when data is linear / Kernel separable
→ Margin maximization is important
→ Only few support vector decides boundary

→ Not use when model needs direct probability output
→ Dataset has heavy overlap → too many support vector
→ Very large Data set.