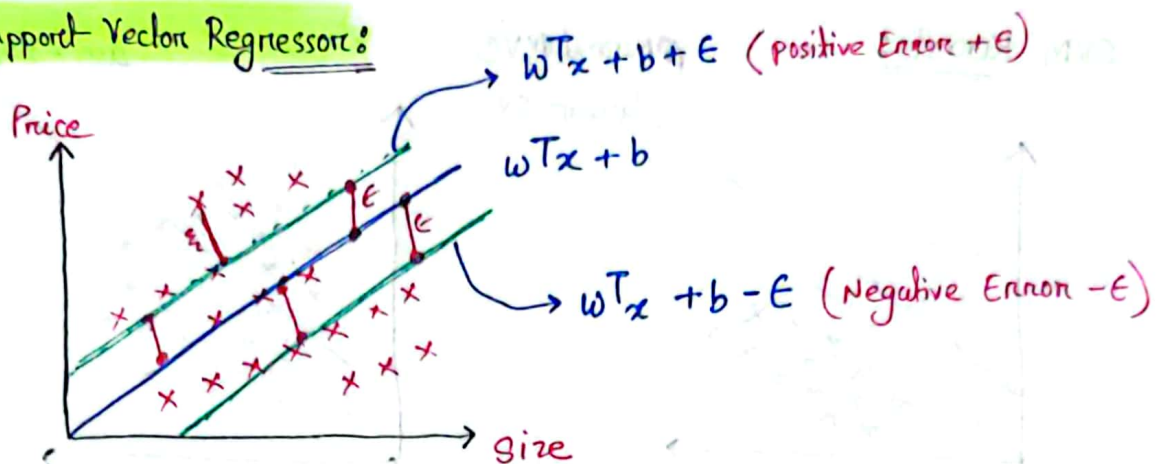


Support Vector Regressor:



Cost Function:

$$\min_{w, b} \frac{\|w\|}{2} + C \sum_{i=1}^n \xi_i \rightarrow \text{hinge loss}$$

$$\text{Constraint} \rightarrow |y_i - w^T x_i| \leq \epsilon + \xi_i$$

Means data points should be under marginal planes
But in real world scenario data points can be outside of marginal points.

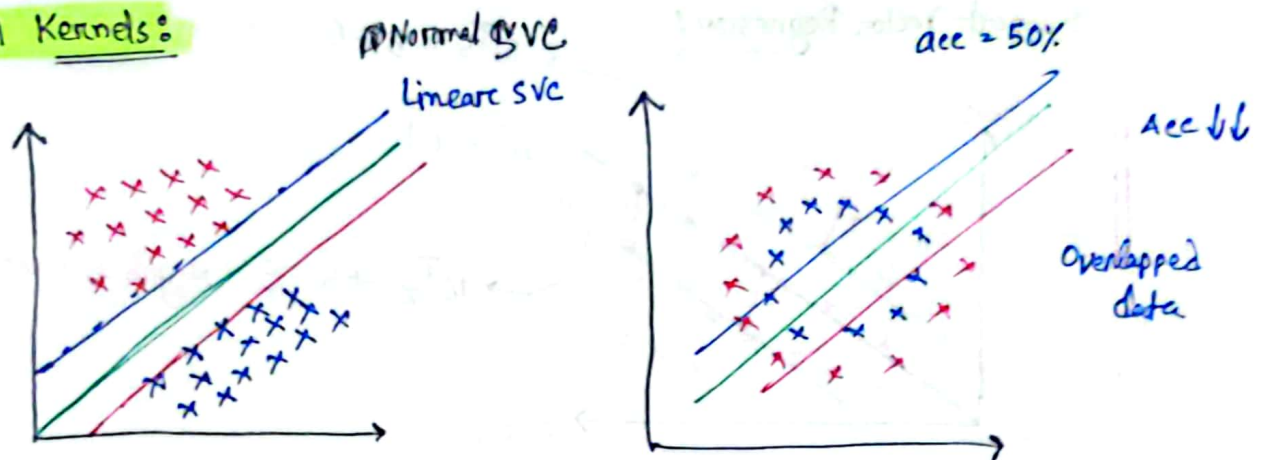


$\xi \rightarrow$ distance of marginal point plane and bFT

$\sum \xi \rightarrow$ Distance of ~~all~~ ^{Sum}

$\xi \rightarrow$ Sum of the distance of the data points with the marginal plane.

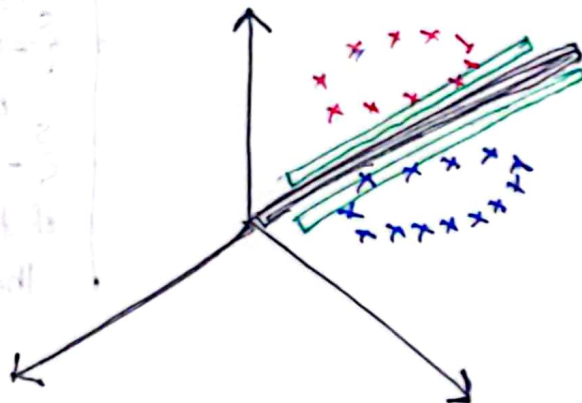
SVM Kernels:



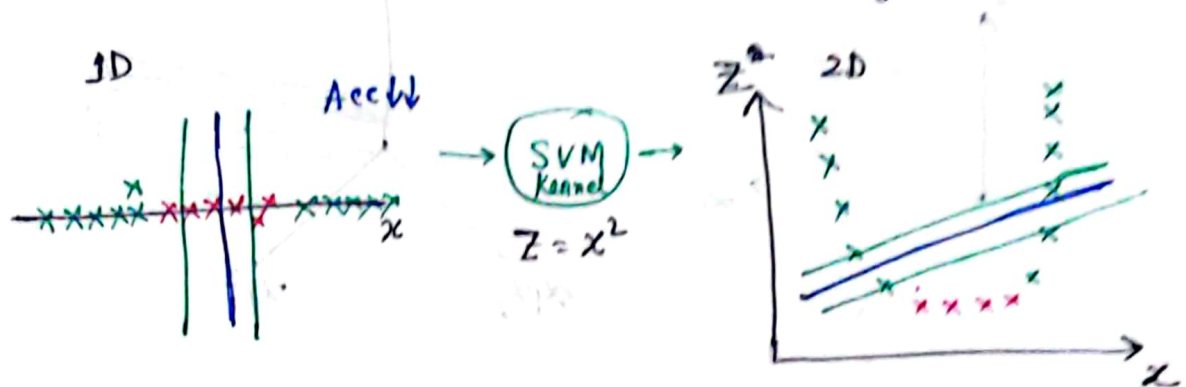
If data orientation is like right chart, the linear SVC won't be able to make the best fit model. It would create low accuracy model. So we have to use 'SVM Kernels' to solve this problem.

What SVM Kernel do?

→ SVM Kernel takes the 2D data, and by applying mathematical formulas it changes the 2D features into 3D features where we can actually separate the overlapped data by marginal hyperplane.



Another example: Suppose data overlapped in a single axis



So, we can see that SVM kernels help data to separate from each other by creating new dimensions.

There are three types of SVM Kernels :

- ① Polynomial Kernel
- ② RBF Kernel
- ③ Sigmoid Kernel

① Polynomial Kernel :

Transformation is happening through this math formula →

$$f(x, y) = (x^T y + c)^d$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

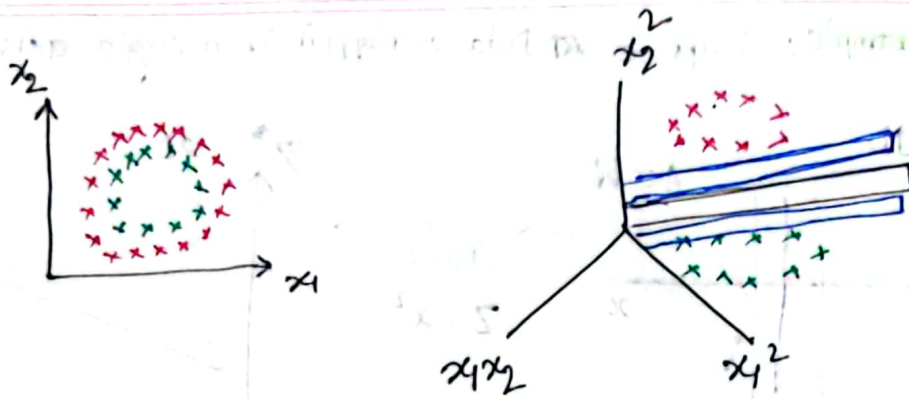
$$= \begin{bmatrix} \underline{x_1^2} & \underline{x_1 x_2} \\ x_1 x_2 & \underline{x_2^2} \end{bmatrix} \quad (3 \text{ unique values})$$

$x = [x_1, x_2]$ (one data point)

$y = [x_1, x_2]$ (another data point)

$c = \text{constant } (1, 2, 3, \dots)$

$d = \text{degree of polynomial}$



RBF Kernel: (Radial Basis Function Kernel)

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$

$\|x - x'\| \rightarrow$ Squared Euclidean distance
 $\sigma =$ free parameter

Sigmoid Kernel:

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

$$\gamma = \frac{1}{2\sigma^2}$$

We actually have to train the model with three of the kernels.

Then we have to check for which kernel, model provides the best accuracy. Then we can choose that kernel.