* Support Vector Machine (SVM)

- Supervised algorithm used for classification and regression.

  SVC: Support Vector Classifier

  SVR: Support Vector Regressor

- We plot each of the data points in n-dimensional space (n = no. of features) with value of each feature being the coordinate of each data point.
- Then we try to find hyperplane which seperates data points for classificaton or try to find hyperplane which has max. no. of data points for regression.
- **Advantages:**
  1. Effective in high dimensional space
  2. Effective if $n >$ no. of samples
  3. Versatile as different Kernels can be used for descision function.
- **Disadvantages:**
  1. If number of features too much high, i.e no. of dimensions high then overfitting occurs, to overcome this we need to choose our Kernel wisely.
  2. Don't provide probability estimate directly and need to use 5 cross validation technique.

A. Building Formula by intution

- Equation of simple line is, (Ref. linear regression)

$$y = mx + c$$

or

$$y = m_1x_1 + m_2x_2 + \dots + m_nx_n + c \quad \text{or}$$

$$h_{\theta_0}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$$

- Algebrically it is same as, (mul. by constants)

$$ax + by + c = 0$$

with,

$$y = \underbrace{\boxed{\frac{-a}{b}}}_{\text{coefficient}} x - \overbrace{\boxed{\frac{c}{b}}}^{\text{intercept}}$$

also for multiple features,

$$ax_1 + bx_2 + dx_3 + \dots + z_n x_n + c = 0$$

to make it bit generalized replace coeff by w_n

$$\boxed{w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n + c = 0}$$

Converting this equation to matrix for ease, where,

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_{n \times 1} \qquad X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}_{1 \times n}$$

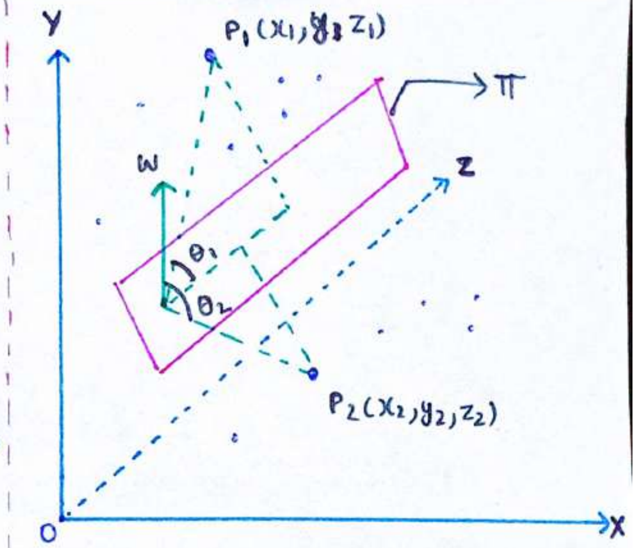
To put up in equation change the order of W by taking Transpose,

$$\boxed{W^T X + c = 0}$$

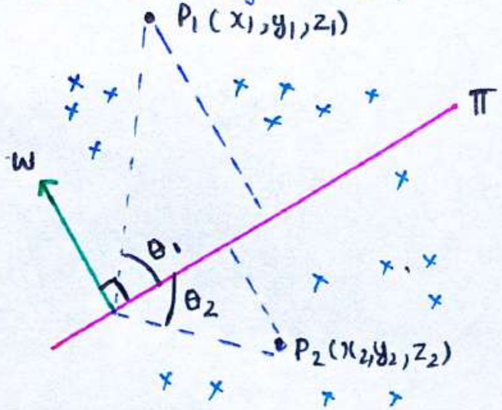
If intercept $c = 0$, $W^T X = 0$ is eq. of line passing via origin.

B. Distance of point from plane

- For simplification, Let we have 3 features, i.e 3 dimensions.
- Our features x_1, x_2, x_3 encoded as x, y, z for geometric intution
- Their are two data points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$



For ease lets bring it to 2D,



- W: unit vector on π ($W \perp \pi$)
- π: a plane
- Points: X
- Distance of P_1 from π is

$$\frac{W^T P_1}{\|W\|}$$

- Distance of P_2 from π is

$$\frac{W^T P_2}{\|W\|}$$