- As $d = 2/\|W\|$ is our margin with respect to a support vector, which is also error. If we inverse it, $\therefore \|W\|/2$

So, we have to maximize our margin $2/\|W\|$ or minimize our loss or error $\|W\|/2$.

- In most of cases data is not linearly seperable by hyperplane and this condition is resolved by Soft margin SVC.

- For classifying under this case we introduce slack variable $(\xi)(x_i)$ in our equation yielding,

$$y_i (W^T X + c) \geq 1 - \xi_i$$

if $\xi_i = 0$, point is correctly classified else
if $\xi_i > 0$, point incorrectly classified

- Incorrect classification means ξ variable is in incorrect dimension.
- ξ_i is basically an error associated with ξ variable

$$\therefore \text{Average error} = \frac{1}{n} \sum_{i=1}^{n} \xi_i$$

- Our objective is to minimize cost function, which is:

$$\boxed{J = \underset{(w,c)}{\text{minimize}} \frac{\|W\|}{2} + c_i \sum_{i=1}^{n} \xi_i}$$

where

c_i: how many points we can ignore for miss classification

ξ_i: summation of incorrect data points from marginal plane

$c_i \sum_{i=1}^{n} \xi_i$ is Hinge Loss function.

$\Rightarrow$ c_i and ξ_i are hyperparameters.

D. Support Vector Regressor

- As a Regressor it triest to fit a best plane which has maximum number of points.
- In this case our marginal planes equation are updated by introducing marginal error (ϵ),

$\therefore$

$$\pi^+ = W^T X + C + \epsilon$$

$$\pi = W^T X + C$$

$$\pi^- = W^T X + C - \epsilon$$

- Thus our cost function updates, insted of $\xi(x_i)$ we use ξ_i (ita of i).

$$\therefore \boxed{J = \underset{(w,c)}{\text{minimize}} \frac{\|W\|}{2} + c_i \sum_{i=1}^{n} \xi_i}$$

- And constraints are;

$$\boxed{|y_i - W^T X| \leq \epsilon + \xi_i}$$

where

ϵ: margin of error (to decide original plane)

ξ_i: error above the margin