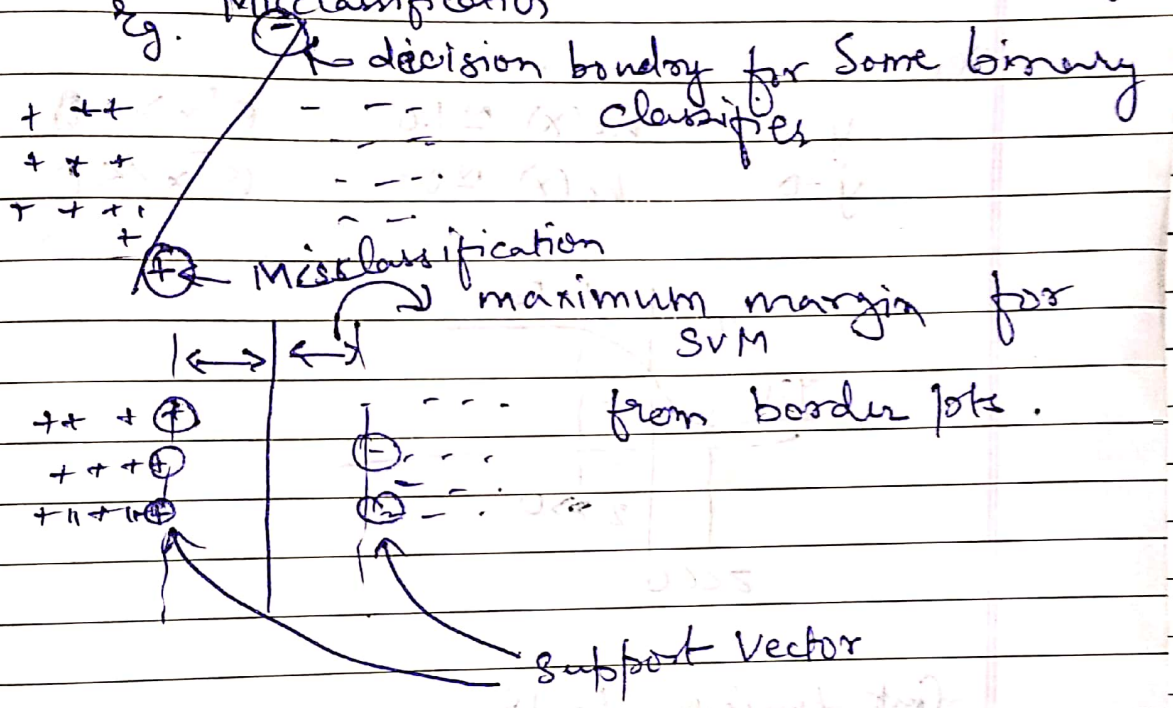
Support Vector Machine

→ well used classifier
→ Binary classifier.
→ Large-margin classifier.

Margin b/w the +ve classes is layer as compared to other binary classifier

Eg. Misclassification



Why SVM should perform better?

data => sample is not proper representation of whole data. if new point comes which at border b'coz first D.B it is misclassified

SVM by choosing Maximum Margin tries to reduce this Misclassification because of improper D.B.

Scanned with CamScanner

Development of SVM from Logistic Regression.

LR ~~$h_\theta(x) = g(\theta^T x)$~~.

$z = \theta^T x$

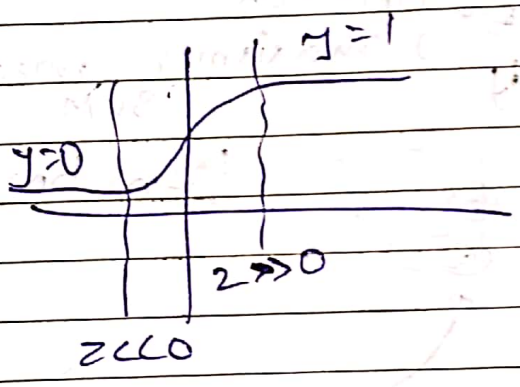
$y = 1 \quad h_\theta(x) \geq 0.5 \quad z \geq 0$

$y = 0 \quad h_\theta(x) < 0.5 \quad z < 0$

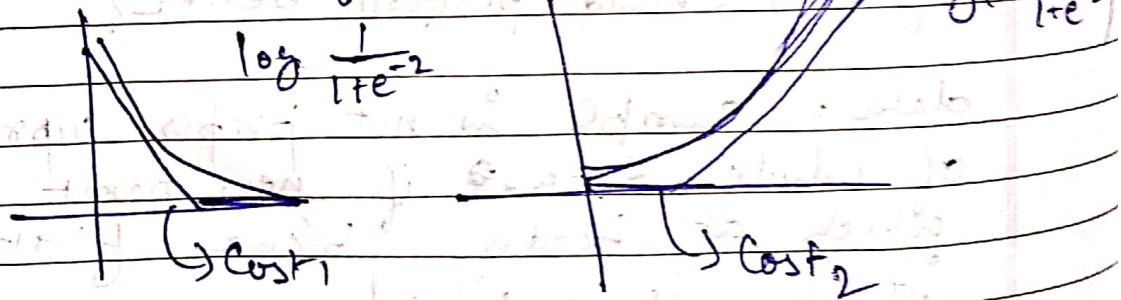
SVM

$y = 1 \quad h_\theta(x) \simeq 1.0 \quad \theta^T x >> 0$

$y = 0 \quad h_\theta(x) \simeq 0.0 \quad \theta^T x << 0$



Cost function for LR



piece wise linearization for SVM

Cost

SVM

$$\min \sum_{i=1}^{m} y^i \text{cost}_1(\theta^T x) + (1-y^i)\,\text{cost}_0(\theta^T x^i) + \frac{\lambda}{2}\sum_{j=0} \theta_j$$

A + λB

B → Regularization

Regularization is done so that there is no overfitting on training data of Machine learning Parameters.

~~min~~ !!! $\min ((u-5)^2 + 1) \times 10$ → $u = 5$

$\min 10(u-5)^2 + 10$ → $u = 5$

Minima remain same when multiplied

$A + \lambda B$.

$\min A + \lambda B = \frac{1}{\lambda} A + B$

$= CA + B$

Scanned with CamScanner