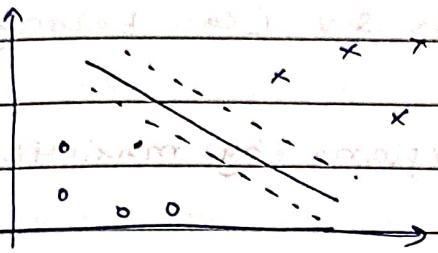
# SVM

(Not very parameters to fiddle with)

* **Visualizing Margin:** The No-mans Band



where we don't find training samples

Margin: width of a band around decision boundary without any training samples.

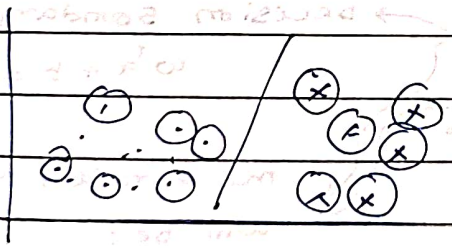
Narrow margin → chances of misclassification ↑

Margin varies with the pos$^n$ & orientation of the seperating hyperplane.

Visualizing Margin: Bubbles around samples.

margin: Radius of a region around each training sample, through which the decision boundary can't pass.

As margin ↑ ; feasible region reduces.



circles influencing the decision boundary.

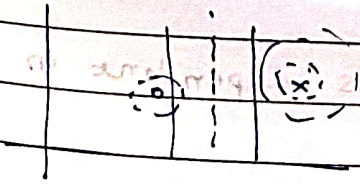
↳ only a few samples controlling decision boundary.

Some samples control the decision boundary.

→ samples that support the decision boundary are called support vectors.

Margin: Band vs Bubbles

Both interpretations yield the same decision boundary



SVM tries to find min distance from nearest support vectors.

Best line SVM: SVM finds the min distance of the frontier from the closest S.V (can belong to any class)

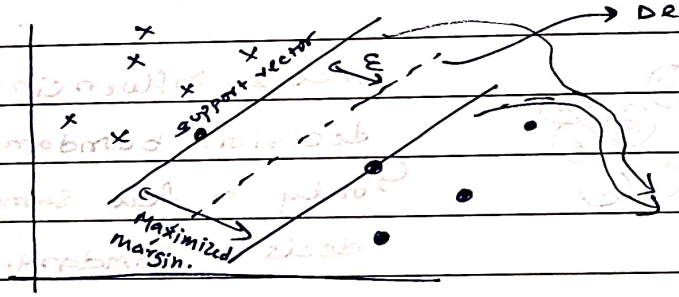
SVM learns the best hyperplane by maximizing the margin (one that lies

Types of SVM:

1. linear SVM: linear SVM is used for linearly seperable data, which means if a dataset can be classified into two classes by using a single straight line.

2. Non-linear SVM:

* Formalizing the Margin:




Decision Boundary:

$w^T x + b = 0$

The parallel hyperplanes will be:

$w^T x + b = \pm \varepsilon$

$$w_i x_i + w_0 = \sum_{i=0}^{N} w_i x_i$$

↳ depending on # of features.

Initial Intuition

In order; to find max value, it may take large values of w.

NOTE: The value of $w^T x_i + b$ is dependent on the scale of w.

## * Maximizing the Margin

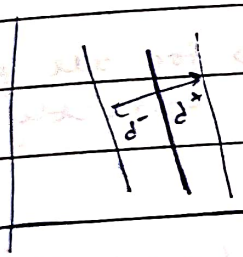
→ we want a classifier (linear seperator) with as big a margin as possible.

→ $(x_0, y_0)$ to a line $Ax + By + C = 0$

$$\Rightarrow \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}$$

→ The dist b/w $H_0$ and $H_1$ is; $\frac{|w^T x + b|}{\|w\|} = \frac{1}{\|w\|}$

→ Total distance b/w $H_1$ & $H_2$ is; $\frac{2}{\|w\|}$




→ can be combined into: $y_i (w^T x_i + b) \geq 1$

* Hinge loss

test sample  decision boundary

How far is your point from classification.

→ The hinge loss incorporates a margin / distance from the classification boundary into the cost calculation.

→ Hinge loss increases linearly.

→ Minimize norm $\|w\|$.

Total margin: $\frac{2}{\|w\|}$

$\frac{1}{2}\|w\|^2 \quad y_i(w^T x - 1) \geq 0$ → for all points correctly classified

2nd part

which samples are misclassified. (outside/inside margin)

$$\frac{1}{n}\sum_{i=1}^{n} \ell(w; (x_i, y_i))$$

Hinge loss Formula

- Hinge loss: Minimize

$$\frac{\lambda}{2}\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n} \ell(w; (x_i, y_i)) \quad (x_i\text{'s are augmented})$$

Samples which are misclassified

where;

$$\ell(w; (x_i, y_i)) = \max\{0, 1 - y_i \langle w, x_i \rangle\}$$

- $\langle w, x_i \rangle$ denotes the inner product of vectors
- Note: For correctly classified samples $y_i \langle w, x_i \rangle \geq 1$
- $\lambda$ → relative importance of margin & training loss.

## * SVM formulation

→ Need to maximize relative margin (not absolute)

- Minimize $|w|$ for mixed margin or maximize margin for $|w| = 1$.
- Training samples should be outside the margin.

→ ie; Minimize $J(w) = \frac{1}{2} w^T w$ ; subject to ,

$$\forall i \; ; \; y_i (w^T x_i + b) \geq 1$$

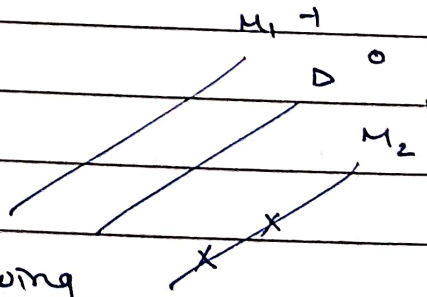
→ This is a constrained quadratic optimization problem.

- Can be solved by the Lagrangian multiplier method
  1. use the method of Lagrange Multipliers ($\alpha$) to modify $J(w)$ to $Q(\alpha)$
  2. Use QP solver to get $\alpha_i$
  3. Use $\alpha$ to find $w$

## * Learning a linear SVM

→ Convert the constrained minimization to an unconstrained optimization problem - represent constraints as penalty terms:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + \text{penalty\_term}$$




→ For data $\{(x_i, y_i)\}_{i=1}^{N}$ ; use the following penalty

$$\max_{\alpha_i \geq 0} \alpha_i \left[1 - y_i (w^T x_i + b)\right] = \begin{cases} 0 & \text{if } y_i (w^T x_i + b) \geq 1 \\ \infty & \text{otherwise.} \end{cases}$$