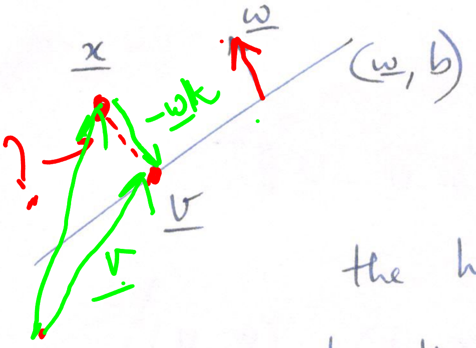
Margin. Distance of a point $\underline{x}$ to a line.

Consider a point $\underline{x}$ in a d-dimensional space.

In the same space there is a hyperplane $(\underline{w}, b)$. — $\underline{w}$: a vector orthogonal

Non homogeneous representation



We want to find out the point on the hyperplane that is closest to $\underline{x}$.

Let $\underline{v}$ be the point for which $\|\underline{x} - \underline{v}\|$ is minimum.

Since $\underline{w}$ is $\perp$ to the hyper plane, we can write $\underline{v} = \underline{x} - k\underline{w}$ where k is a scalar.

Since $\underline{v}$ lies on the hyperplane, we have

$\langle \underline{w}, \underline{x} \rangle + b = 0$

$$\langle \underline{w}, \underline{v} \rangle + b = 0$$

$\underline{v}$ is on the line.

$$\therefore \langle \underline{w}, \underline{x} - k\underline{w} \rangle + b = 0$$

$$\langle \underline{w}, \underline{x} \rangle - k \langle \underline{w}, \underline{w} \rangle + b = 0$$

$$\langle \underline{w}, \underline{x} \rangle - k \|\underline{w}\|^2 + b = 0$$

We assume $\|\underline{w}\|^2 = 1$

$$\therefore \langle \underline{w}, \underline{x} \rangle - k + b = 0$$

$$\therefore k = \langle \underline{w}, \underline{x} \rangle + b$$

scalar

$\underline{x} - \underline{v} = k\underline{w}$

$$\therefore \underline{v} = \underline{x} - \left(\langle \underline{w}, \underline{x} \rangle + b \right) \underline{w}$$

We have ensured that $\underline{v}$ lies on the hyperplane.

But is it the closest vector to $\underline{x}$?

Let us check.

Take any point $\underline{u}$ on the hyperplane.

$$\langle \underline{w}, \underline{u} \rangle + b = 0$$

Now consider $\|\underline{x} - \underline{u}\|^2 = \|\underline{x} - \underline{v} + \underline{v} - \underline{u}\|^2$

$$= \|\underline{x} - \underline{v}\|^2 + \underbrace{\|\underline{v} - \underline{u}\|^2}_{\text{always positive}} + 2\langle (\underline{x} - \underline{v}), (\underline{v} - \underline{u}) \rangle$$

$$\geq \|\underline{x} - \underline{v}\|^2 + 2\langle \underbrace{(\underline{x} - \underline{v})}_{\text{substitute}}, (\underline{v} - \underline{u}) \rangle$$

$$= \|\underline{x} - \underline{v}\|^2 + 2\langle (\underbrace{\langle \underline{w}, \underline{x} \rangle + b}_{k}) \underline{w}, (\underline{v} - \underline{u}) \rangle$$

$$= \|\underline{x} - \underline{v}\|^2 + 2\underbrace{(\langle \underline{w}, \underline{x} \rangle + b)}_{\text{scalar}} \langle \underline{w}, (\underline{v} - \underline{u}) \rangle$$

$$= \|\underline{x} - \underline{v}\|^2 + 2(\langle \underline{w}, \underline{x} \rangle + b)\underbrace{(\overset{-b}{\langle \underline{w}, \underline{v} \rangle} - \overset{-b}{\langle \underline{w}, \underline{u} \rangle})}_{=0}$$

$$= \|\underline{x} - \underline{v}\|^2$$

$= 0$ because both $\underline{u}$ and $\underline{v}$ are points on the hyperplane.

$$\langle \underline{w}, \underline{v} \rangle + b = 0$$
$$\langle \underline{w}, \underline{u} \rangle + b = 0$$

We have showed that $\boxed{\|\underline{x} - \underline{u}\|^2 > \|\underline{x} - \underline{v}\|^2}$

So $\underline{v}$ is the point on the hyperplane and is closest to $\underline{x}$. Therefore, the shortest distance is $\boxed{|\langle \underline{w}, \underline{x} \rangle + b|}$ if $\boxed{\|w\| = 1}$

In SVM. → otherwise $\dfrac{|\langle \underline{w}, \underline{x} \rangle + b|}{\|\underline{w}\|}$

Halfspaces not robust: Support Vector Machine (SVM)

Can have poor generalization

Given a linearly separable training data, a half space classifier ensures

correct classification

$$y_i(\langle \underline{w}, \underline{x_i} \rangle + b) > 0 \quad \forall\, i = 1 \text{ to } m,$$

There can be several solutions for $\underline{w}$ that satisfy all constraints.

The ERM learning rule says pick any candidate solution

Which of these solutions will generalize better?

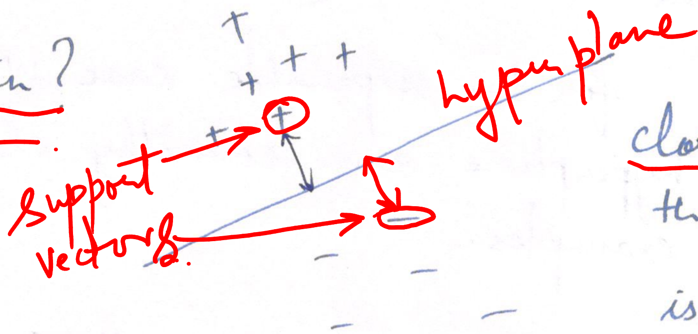
Goal of ML.

What is the general learning rule which can ensure that the selected hyperplane generalizes well?

Rule: Choose the hyperplane with the maximum margin.

What is margin?

distance ??



Distance of the closest point to the hyperplane is the margin.

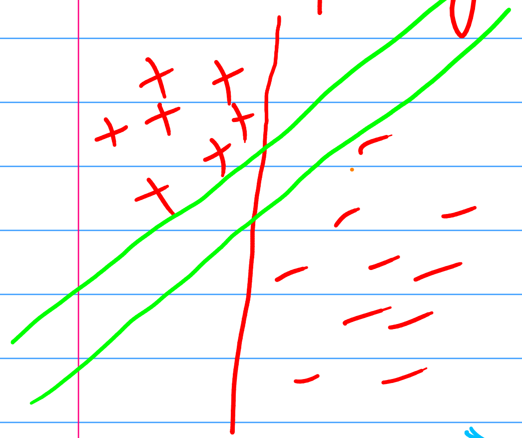
Large margin is robust to perturbation of instances.

Hard SVM Learning Rule:

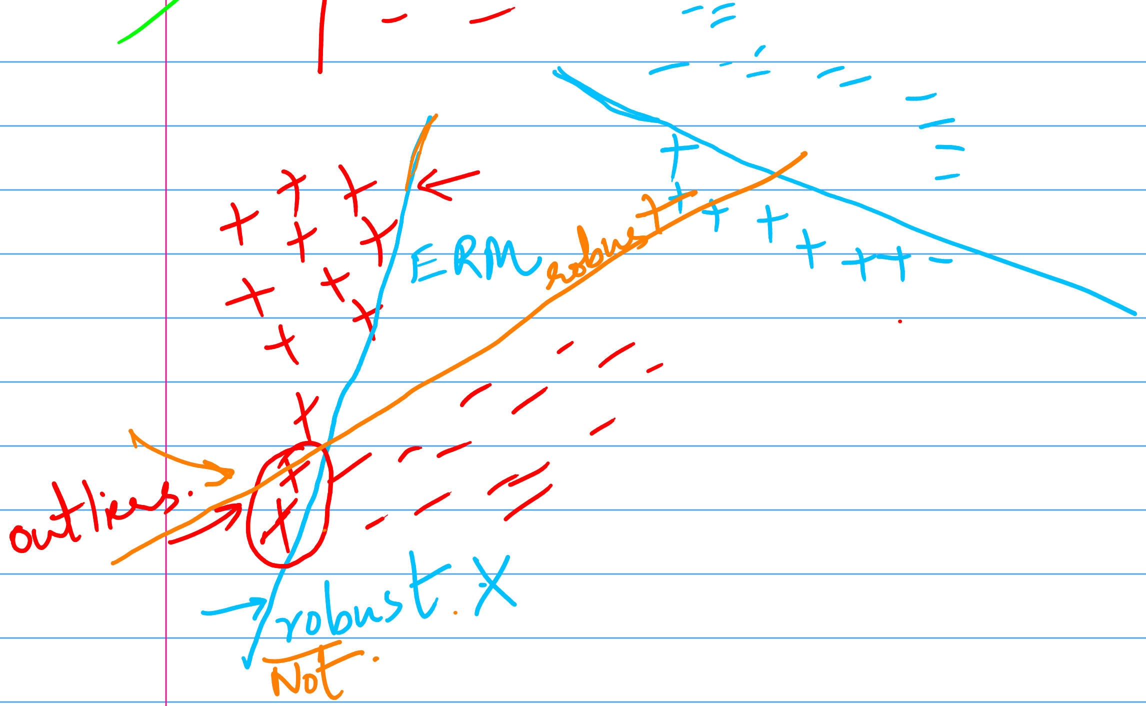
Support Vector Machine

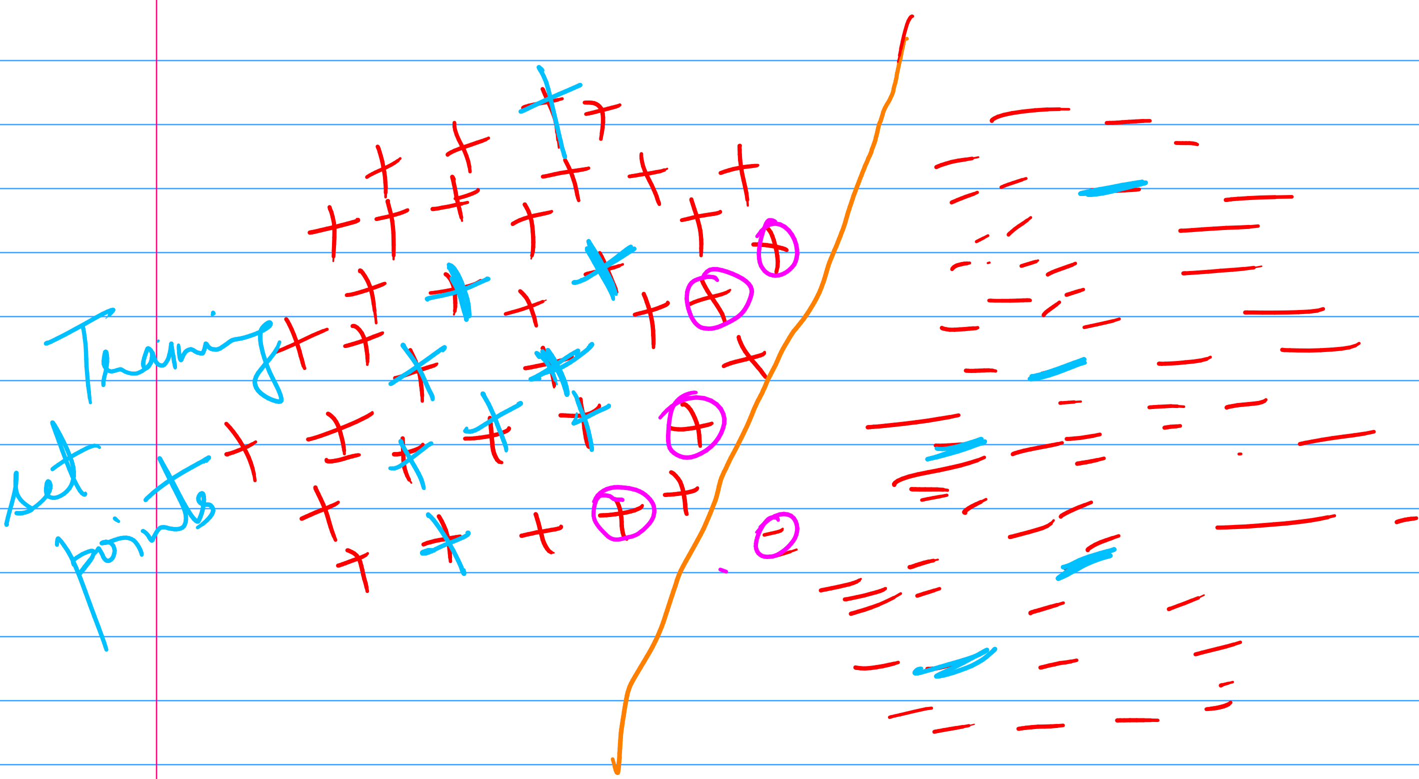
Choose the ERM hyperplane that separates the training set with the largest possible margin.

Overfitting → learning the noise



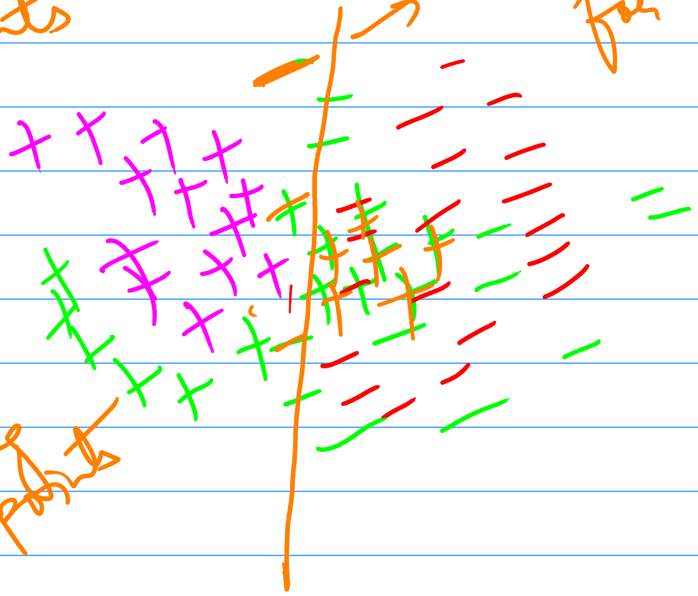
$\underbrace{\text{ERM learning.}}_{\text{weak.}}$

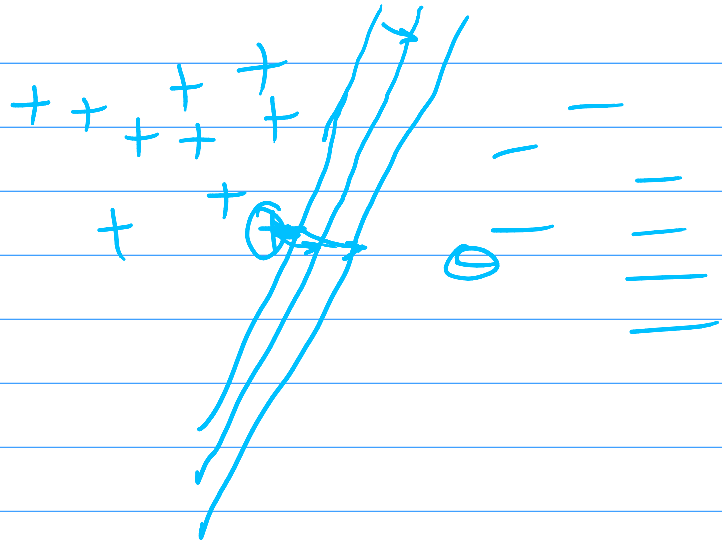




Perturbation → test points can be different from the training points

→ 100% correct classification for training set





Hard: ⇒ insisting on all points being correctly classified.

dataset is separable

all constraints being satisfied.

$\begin{bmatrix} a \\ b \end{bmatrix}$. $ax + by + c = 0$

The distance of a training point x_i to a hyperplane $(\underline{w}, b)$ is

$$\frac{|\langle \underline{w}, \underline{x_i} \rangle + b|}{\|\underline{w}\|}$$

Distance.

Assuming that $\|\underline{w}\| = 1$, the distance of the closest point to the hyperplane is

$$\min_{i \in [m]} |\langle \underline{w}, x_i \rangle + b|$$

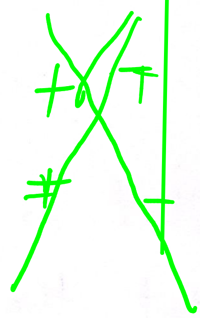
Margin

The hard SVM learning rule says that pick up the hyperplane which maximizes the margin.

learned.

separating hyperplane

$$\underset{\substack{(\underline{w}, b) \\ \|\underline{w}\| = 1}}{\arg\max} \underbrace{\left[\min_{i \in [m]} |\langle \underline{w}, x_i \rangle + b| \right]}_{\text{margin}}$$



such that $\forall i \quad \boxed{y_i(\langle \underline{w}, x_i \rangle + b) > 1}$

these are m (#examples)

for the separable case when the hyperplane can correctly classify all the examples.

Hard margin SVM learning rule is applicable to only separable data.

For the separable case we are sure that

$$y_i(\langle \underline{w}, x_i \rangle + b) > 0$$

$$\therefore \quad |\langle \underline{w}, x_i \rangle + b| = y_i(\langle \underline{w}, x_i \rangle + b)$$

$\therefore$ the equivalent problem is $\underset{\substack{(\underline{w}, b) \\ \|\underline{w}\| = 1}}{\arg\max} \; \min_{i \in [m]} y_i(\langle \underline{w}, x_i \rangle + b)$