CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Linear Machines & SVM

Arizona State University

Table of contents

Quick Recap - Logistic Regression

- Data - X (features) and Y (labels)
- Learn P(Y|X) assuming a logistic function — Discriminative model.
- $g(x) = w^T x$ is called the linear discriminant function.

w

Linear Discriminant functions

- Two types of notations-

$$g(x) = w^T x \qquad \text{or} \qquad g(x) = w^T x + w_0$$

$w \quad w_0$

weights

bias | threshold

- For 2-class problem, learn **w** such that -

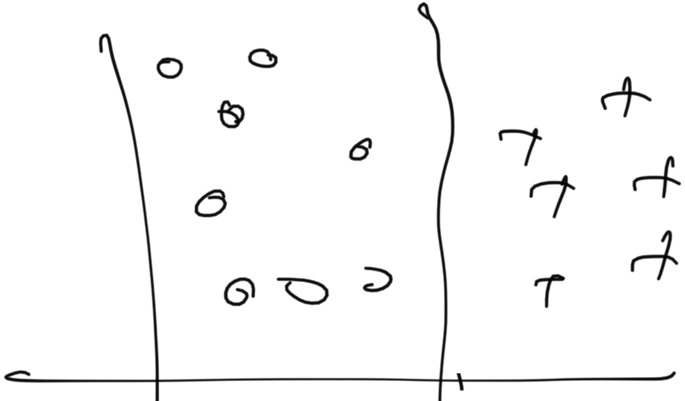
$$w^T x > 0 \rightarrow \text{Class } 1$$

$$w^T x \leq 0 \rightarrow \text{class } 0.$$

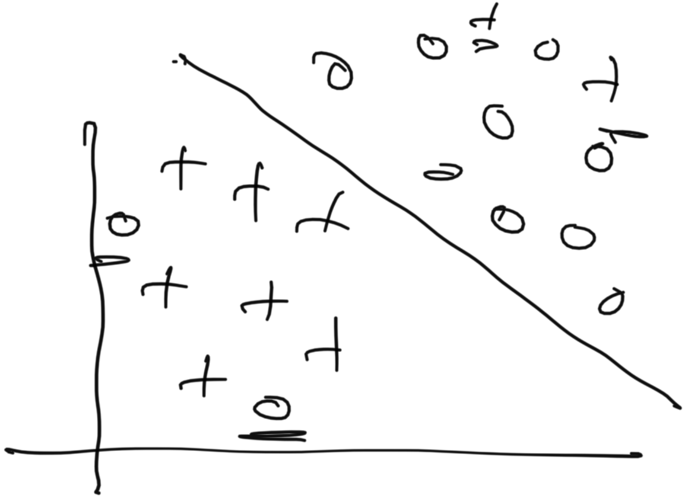
What is Linear Separability?

$w^T x$

- If there is atleast one solution of **w** such that g(x) classifies all the samples in the training data, then the data is said to be linearly separable.

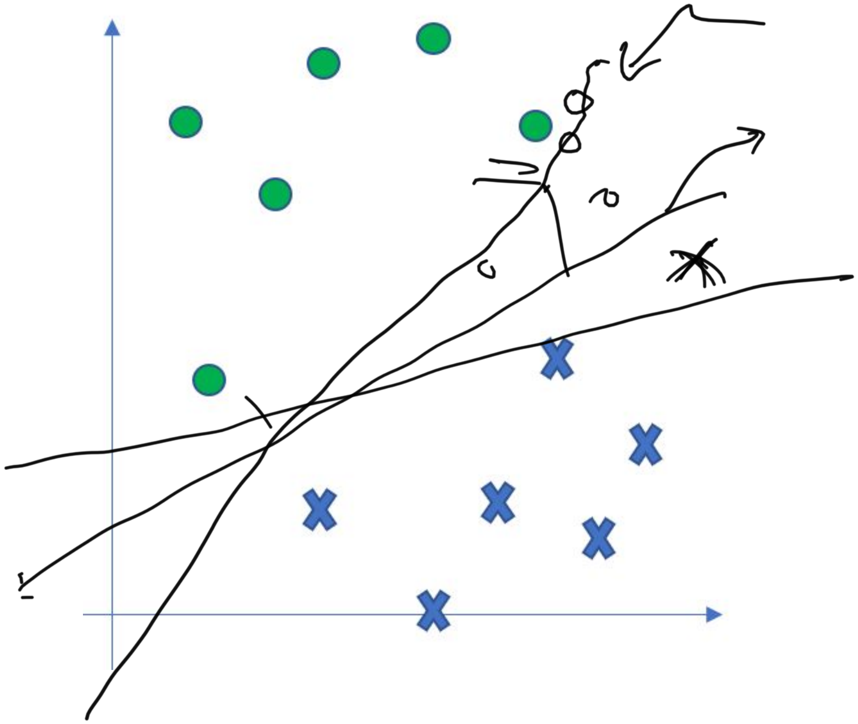


Linearly separable data

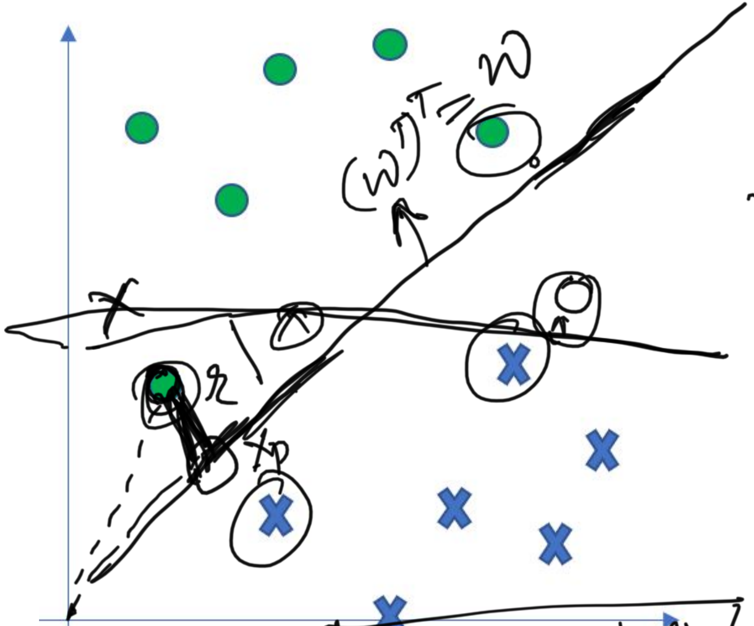


Non-linear separable data

Is the solution unique?



Concept of Margin



$$g(x) = w^T x + w_0 = 0$$

$$r\frac{w}{\|w\|}$$

$$x = x_p + r\frac{w}{\|w\|}$$

$$x_p = x - r\frac{w}{\|w\|} \quad (1)$$

$$g(x_p) = 0$$

$$w^T x_p + w_0 = 0$$

$$w^T\left(x - r\frac{w}{\|w\|}\right) + w_0 = 0$$

$$w^T x - r\frac{w^T w}{\|w\|} + w_0 = 0$$

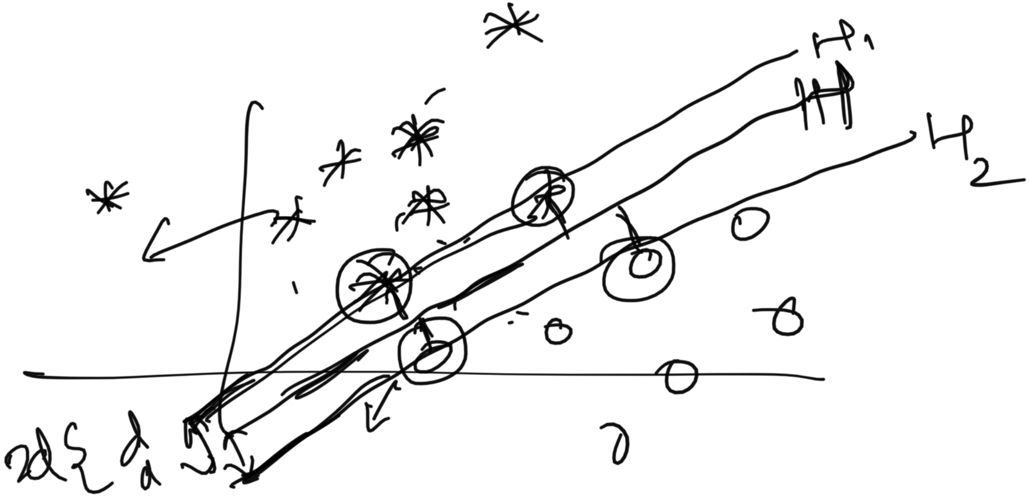
$$\boxed{w^T x + w_0 = r\|w\|}$$

$$g(x) = r\|w\|$$

$$\boxed{r = \frac{g(x)}{\|w\|}}$$

Concept of Margin

- Margin of a sample x (w.r.t. the decision plane) is defined as the distance from x to the plane.
- For a classifier, the margin should be as large as possible for better performance.

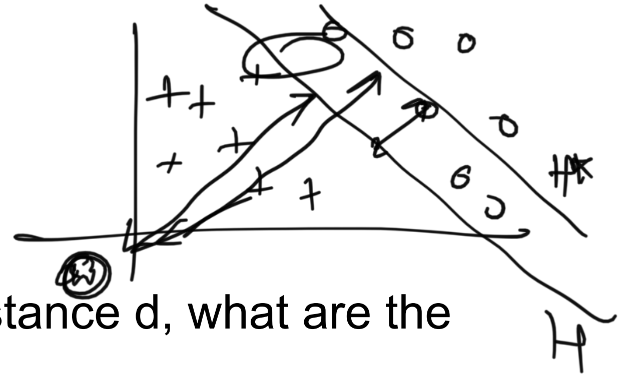


Support Vector Machines (SVM)

→ Maximum margin classifier.

- Key idea - To find the decision boundary such that the margin is maximized.
- Data - $<x^{(i)}, y^{(i)}>$, $y^{(i)} \in \{-1,1\}$, $x^{(i)} \in \boldsymbol{R}^d$, for all $i=1,\ldots,n$
- Assume that the data is linearly separable

SVM - Problem Formulation



- Given separating plane H: $w^t x + b = 0$ and distance d, what are the equations for H_1 and H_2?
- Consider H* plane with equation - $H^* = w^T x + b - \|w\| d$

$$d(\text{origin}, H) = \frac{g(0)}{\|w\|} = \frac{b}{\|w\|}$$

$$d(\text{origin}, H^*) = \frac{g^*(0)}{\|w\|} = \frac{b - \|w\| d}{\|w\|}$$

$$d(H, H^*) = d$$

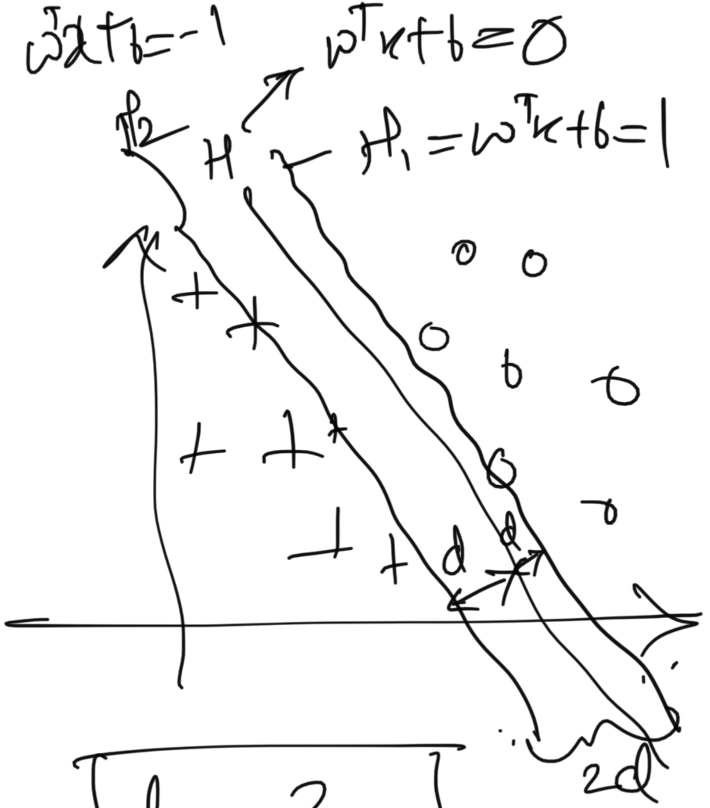
SVM - Problem Formulation

$$H_1 = w^T x + b = \|w\| d$$

$$H_2 = w^T x + b = -\|w\| d$$

$$H = w^T x + b = 0$$

$$\left\{\begin{aligned} H_1 &= w^T x + b = 1 \\ H_2 &= w^T x + b = -1 \\ H &= w^T x + b = 0 \end{aligned}\right\} \rightarrow \boxed{d = \frac{2}{\|w\|}}$$



SVM - Problem Formulation

- Plane equations-

H_1

H_2

H

- Margin -

$$d = \frac{2}{\|w\|}$$

$$d = \frac{2}{\|w\|}$$

SVM - Problem Formulation

$$\{\mathbf{w}^*, b^*\} = \underset{\mathbf{w},b}{\operatorname{argmin}} \|\mathbf{w}\| \;\; or \;\; \{\mathbf{w}^*, b^*\} = \underset{\mathbf{w},b}{\operatorname{argmin}} \tfrac{1}{2}\|\mathbf{w}\|^2$$

Subject to

$$\left.\begin{aligned} \mathbf{w}^t\mathbf{x}^{(i)} + b \geq 1 \quad \text{for } y^{(i)} = +1 \\ \mathbf{w}^t\mathbf{x}^{(i)} + b \leq -1 \quad \text{for } y^{(i)} = -1 \end{aligned}\right\}$$

The constraints can be combined into:

$$y^{(i)}(\mathbf{w}^t\mathbf{x}^{(i)} + b) - 1 \geq 0 \quad \forall i$$

Questions?