Support Vector Machines

HW#2 is due 02/17

→ ionosphere.csv

Our goal in this lecture is to revisit linear classification for two-class problems.

$\hat{y} = \text{sign}(g(x))$

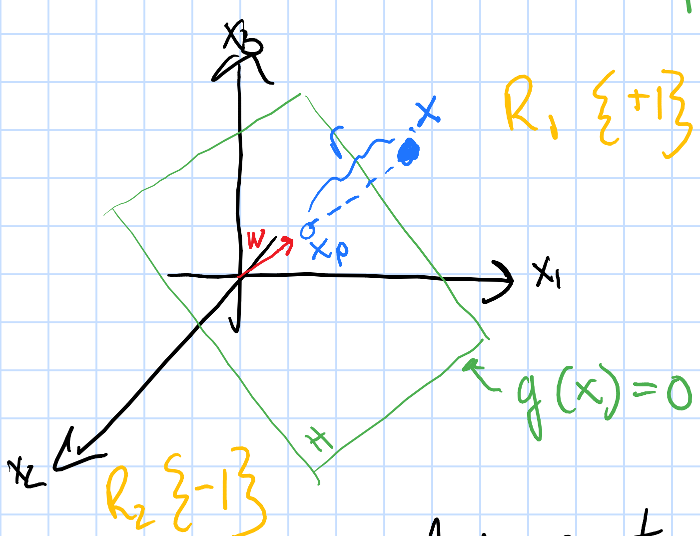
Notation

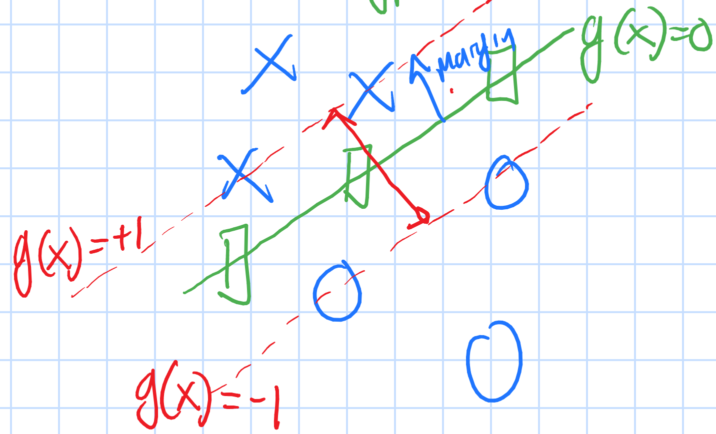
$x \in \mathbb{R}^D$

$y \in \{\pm 1\}$

$g(x) = w^T x + b \Rightarrow w \in \mathbb{R}^D, b \in \mathbb{R}$

The discriminant function gives us an algebraic measure of the distance from a point, x, to a hyperplane.





Any point in the $\mathbb{R}^D$ space can be written as:

$$x = x_p + r\frac{w}{\|w\|_2}$$

x_p: projection of x onto the hyperplane

r: measure from the hyperplane

What if $r = 0$. [on the plane]

$r > 0$ [$y = +1$]

$r < 0$ [$y = -1$]

Substitute x into $g(x)$

$$g(x) = w^T x + b$$

$$= w^T \left(x_p + r \frac{w}{\|w\|_2} \right) + b$$

$$= \underbrace{w^T x_p + b}_{g(x_p) = 0} + r \frac{w^T w}{\|w\|_2}$$

$w^T w = \|w\|_2^2$

$$\Rightarrow \boxed{r = \frac{g(x)}{\|w\|_2}}$$

margin

x_1 is where $g(x) = 1$

$r_1 = \frac{1}{\|w\|_2}$

x_2 is where $g(x_2) = -1$

$r_2 = \frac{-1}{\|w\|_2}$

Back to the prediction task. Goal: Maximize the margin

$y_i = +1 \qquad w^T x_i + b \geq 1$

$y_i = -1 \qquad w^T x_i + b \leq -1$

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

$\arg\min_z f(z)$

s.t. $h(z) = 0$

$g(z) \leq 0$

$L(z, \lambda) = f(z) + \lambda h(z)$

Considering the points that lie on the margin, the best hyperplane is the one that

$$\arg\max_{w,b} \left\{ \frac{2}{\|w\|_2} \right\} \quad (1)$$

$$\text{s.t.} \quad y_i (w^T x_i + b) \geq 1 \quad \forall i$$

For (1) and (2), the data are perfectly seperable.

Rewrite as (2)

$$\arg\min \frac{1}{2} \|w\|_2^2, \quad \text{s.t.} \quad y_i (w^T x_i + b) \geq 1 \quad \forall i$$

We need to form the Lagrangian function to solve the constrained optimization task.

$$\boxed{L(w, b, \alpha)} = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i \left[y_i (w^T x_i + b) - 1 \right]$$

$$* \quad \alpha_i \geq 0 \quad \forall i$$

Find $\frac{\partial L}{\partial w}$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \implies \boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

Find $\frac{\partial L}{\partial b}$

Another constraint!!

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0 \implies \boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$$