Separating Hyperplanes (support vector machine)

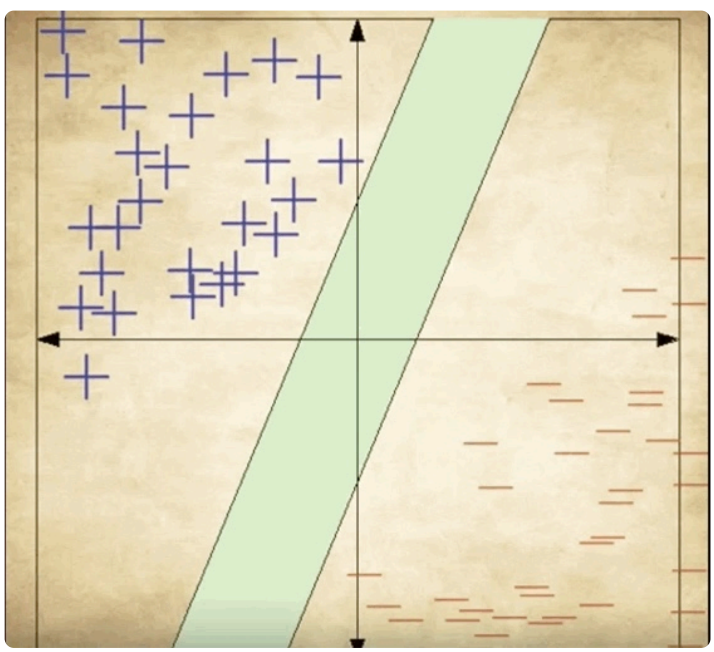
LDA /QDA /NB $p(x)$ and $p(Y|X)$

Logistic Regression $p(Y|X)$

Support Vector Machine $w^Tx+b\geq 0$ directly estimate $\theta=\begin{bmatrix}w\\b\end{bmatrix}$

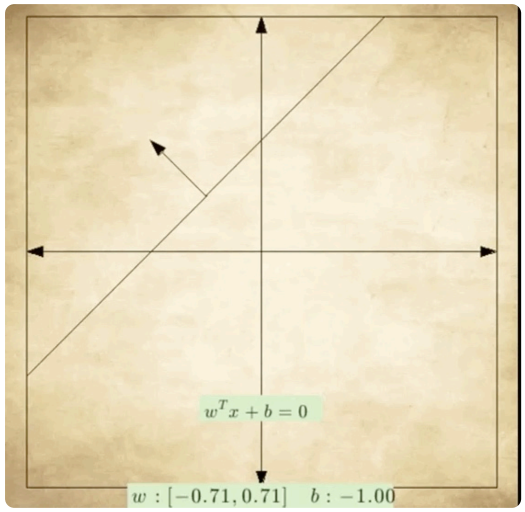
Vapnik's Principle:

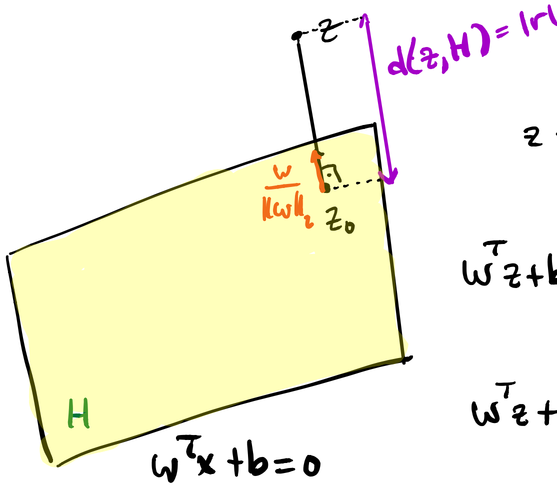
> "When solving a problem of interest, do not solve a more general problem as an intermediate step."
>
> "Don't solve a harder problem than you have to"



Hyperplane:

$$H = \{ x : w^T x + b = 0 \}$$





Unit norm normal vector

$$z = z_0 + \frac{w}{\|w\|_2} \cdot r$$

$$\Downarrow$$

$$w^T z + b = \underbrace{w^T z_0 + b}_{0} + \|w\|_2 \cdot r$$

$$w^T z + b = \|w\|_2 \cdot r$$

distance btw z and H : $d(z,H) = |r| = \dfrac{|w^T z + b|}{\|w\|_2}$

Let $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ be training data $y_i \in \{-1, 1\}$

Data is linearly separable if $\exists w, \exists b$

$$y_i (w^T x_i + b) \geq 0 \quad \forall i = 1, \ldots, n$$

$H = \{x : w^T x + b = 0\}$ is a separating hyperplane

Maximum Margin Hyperplane

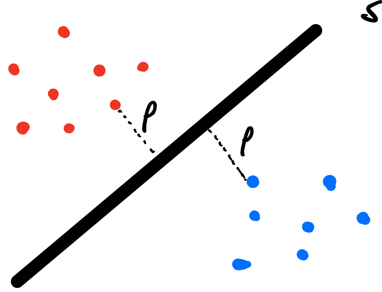
Margin ρ of a sep. hyperplane is

$$\rho(w,b) = \min_{i=1,\dots,n} d(x_i, H) = \min_{i=1,\dots,n} \frac{|w^T x_i + b|}{\|w\|_2}$$

Max. margin separating hyperplane is the solution of

$$\max_{w,b} \rho(w,b)$$

$$\text{s.t. } y_i(w^T x_i + b) \geq 0 \quad \forall i$$



not unique $(\alpha w, \alpha b)$ gives the same hyperplane

Scale w and b by $\dfrac{1}{\min_{i=1,\dots,n} |w^T x_i + b|}$.

Now $\rho(w,b) = \dfrac{1}{\|w\|_2}$

$$\max_{w,b} \frac{1}{\|w\|_2} \qquad = \qquad \max_{w,b} \frac{1}{\|w\|_2}$$

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

$y_i(w^T x_i + b) = 1$ for some i ← this constraint is automatically satisfied (we can drop it)

$$\min_{w,b} \ \frac{1}{2}\|w\|_2^2$$

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

Hard-Margin Support Vector Machine

Constrained Optimization Problem (Quadratic Optimization with linear constraints).

$$\min \ \frac{1}{2}\|w\|_2^2 + C \cdot \frac{1}{n}\sum_{i=1}^{n} s_i$$

$$y_i(w^T x_i + b) \geq 1 - s_i \quad \forall i$$

$$s_i \geq 0 \quad \forall i$$ (slack variables)

Soft-Margin Support Vector Machine

C is a tuning parameter