LINEAR CLASSIFIER

H = Space of hypotheses.
= Half-spaces or Linear Separators.

$$\theta_1 x_{i1} + \theta_2 x_{i2} + \cdots + \theta_d x_{id} \lessgtr b$$

$$\text{sign}(\vec{\theta}\cdot\vec{x}_i - b) = l_i, \quad l_i \in \{+1, -1\} \qquad i = 1, \ldots, n.$$

Projective Space $\Rightarrow$ $\vec{w} = (\vec{\theta}, -b)$

$$\begin{cases} \vec{a}_i = (\vec{x}_i, +1) & \text{if } l_i = +1 \\ \vec{a}_i = (-\vec{x}_i, -1) & \text{if } l_i = -1 \end{cases}$$

$$\text{sign}(\vec{w}\cdot\vec{a}_i) = \begin{cases} \text{sign}(\vec{\theta}\cdot\vec{x}_i - b) = 1 & \text{if } l_i = +1 \\ \text{or} \begin{cases} \text{sign}(-\vec{\theta}\cdot\vec{x}_i + b) = -\text{sign}(\vec{\theta}\cdot\vec{x}_i - b) \\ = 1 \quad \text{if } l_i = -1. \end{cases} \end{cases}$$

◇ Given a sample

$$S' = \{(\vec{x}_1, l_1), (\vec{x}_2, l_2), \ldots, (\vec{x}_n, l_n)\} \subseteq \mathbb{R}^d \times \{+1, -1\}$$

Rewrite

$$S = \{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\} \subseteq \mathbb{R}^{d+1}, \quad |\vec{a}_i| = 1$$

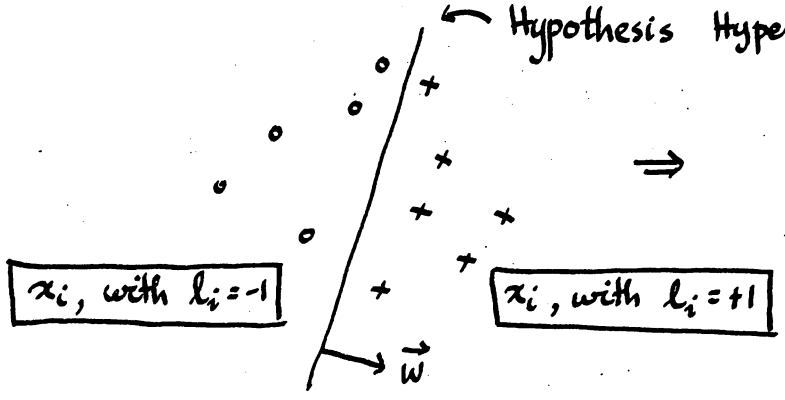
Find a solution vector $\vec{\omega} \in \mathbb{R}^{d+1}$, $|\vec{\omega}| = 1$ such that

$$\forall_{i \in \{1, n\}} \quad \boxed{\vec{\omega}\cdot\vec{a}_i > 0} \qquad \square.$$

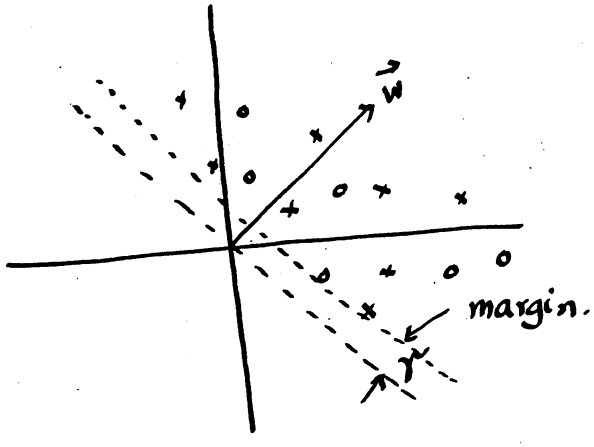
OR SOLVE THE FOLLOWING OPTIMIZATION PROBLEM.

$$\max_{\omega^*} \min_{\vec{a}} \frac{\vec{\omega}^*\cdot\vec{a}_i}{|\vec{\omega}^*|\,|\vec{a}_i|} = \text{margin}$$

CHECK IF THE MARGIN > 0.



Transformed problem.



$$\gamma_P = \max_{\omega^*} \min_i \frac{\vec{\omega}^* \cdot \vec{a}_i}{\|\vec{\omega}^*\|_2 \|\vec{a}_i\|_2}$$

$$\gamma_W = \max_{\omega^*} \min_i \frac{\vec{\omega}^* \cdot \vec{a}_i}{\|\vec{\omega}^*\|_1 \|\vec{a}_i\|_\infty}$$

Check $\gamma_P > 0$ (iff) $\gamma_W > 0$.

Related by some function of d.
d = const.

MULTIPLICATIVE WEIGHTING.

Simulate a repeated game
$\omega^* \approx$ Mixed Strategy (row player)
} WINNOW ALGORITHM

$$M(i,j) = -a_{ij}$$

Problem 1. $w_i \geq 0$, weight must be positive

$$S = \{ (\vec{a}_1, -\vec{a}_1), \cdots, (\vec{a}_n, -\vec{a}_n) \} \subseteq \mathbb{R}^{2(d+1)}$$

$$\vec{w} \cdot \vec{a}_i = w_1 a_{i1} + w_2 a_{i2} + \cdots + w_{d+1} a_{i(d+1)}$$
$$= (w_1^+ - w_1^-) a_{i1} + \cdots + (w_{d+1}^+ - w_{d+1}^-) a_{i(d+1)}$$

$$\vec{\omega} \cdot \vec{a}_i = \langle w_1^+, w_2^+, \dots, w_{d+1}^+, w_1^-, \dots, w_{d+1}^- \rangle \cdot \langle \vec{a}_i, -\vec{a}_i \rangle$$

The Algorithm

v_i = weight in the direction of the i^{th} feature

$v_i^0 = 1 \quad \forall i = 1, \dots, d+1$

Round t

$$\sigma_{i,t} = \frac{v_i^{t-1}}{\sum v_i^{t-1}}$$

$$\omega^t = \langle \sigma_{1,t}, \dots, \sigma_{d+1,t} \rangle$$

Check $\exists_{j_t} \ \vec{\omega}^t \cdot \vec{a}_{j_t} \leq 0 \quad [a_j = \text{misclassified}]$

If ~~$\exists_{j_t}$~~ $\vec{\omega}^t \ a_{j_t}$

If $\forall_j \ \vec{\omega}^t \cdot \vec{a}_j > 0 \Rightarrow$ Success.

Else

$$v_i^t = v_i^{t-1}(1-\varepsilon)^{-a_{i,j_t}} \approx v_i^{t-1}(1+\varepsilon a_{i,j_t})$$

(Neglect $o(\varepsilon)$ terms)

(A) $$\sum_t M(\sigma_{r,t}, \sigma_{c,t}) = \sum_t \sum_i \sigma_{i,t} M(i, j_t) = -\sum_t \sum_i \sigma_{it} a_{i,j_t}$$

$$= -\sum_t \vec{\omega}^t \cdot a_{j_t} \geq 0$$

(B) $$\sum_t M(\sigma_{r,t}, \sigma_{c,t}) \leq \frac{\ln n}{\varepsilon} + (1+\varepsilon) \sum_t M(\sigma_r^*, \sigma_{c,t})$$

$$= \frac{\ln n}{\varepsilon} + (1+\varepsilon) \sum_t -\omega^* \cdot a_{j_t}$$

$$\leq \frac{\ln n}{\varepsilon} + \varepsilon T - r_\omega T$$

Take $\varepsilon = \frac{3r_w}{4}$ (ε Needs to be searched by binary search) $\Rightarrow$ $\frac{r_w}{4} \leq \varepsilon \leq \frac{r_w}{2}$

$$\therefore \quad \frac{4\ln n}{3r_w} + \frac{3r_w}{4} T - r_w T \geq 0$$

$$\frac{r_w T}{4} \leq \frac{4\ln n}{3r_w} \Rightarrow T \leq \frac{16}{3} \frac{\ln n}{r_w^2}$$ (More correct estimate takes into account Binary Search)

$$T = O\left(\frac{\ln n}{r_w^2}\right)$$

A HEURISTIC APPROACH → PERCEPTRON

Additive Update → Instead of Multiplicative.

a) SIMPLE
b) No need to worry about $w_i^t \geq 0$

The Algorithm.

$w^0 \leftarrow (0, 0, \cdots, 0)$

Round t.

Check $\exists_{j_t} \ \vec{w}^t \cdot \vec{a}_{j_t} \leq 0$ [$\vec{a}_{j_t}$ = misclassified]

If $\forall_j \ \vec{w}^t \cdot \vec{a}_j > 0 \Rightarrow$ Success.

Else

$$\vec{w}^{t+1} = \vec{w}^t + \vec{a}_{j_t}$$

Note $\vec{w}^{t+1} \cdot \vec{a}_{j_t} = \vec{w}^t \cdot \vec{a}_{j_t} + \vec{a}_{j_t} \cdot \vec{a}_{j_t}$

$$= \vec{w}^t \cdot \vec{a}_{j_t} + 1 > \vec{w}^t \cdot \vec{a}_{j_t}$$

However, there is no guarrantee that it will not fail for some sample point that was correctly classified so far!

Let's define a potential function

$$\Phi_t = \text{cosine}(\vec{w}^t, \vec{w}^*) = \frac{\vec{w}^t \cdot \vec{w}^*}{\|\vec{w}^t\|_2 \|\vec{w}^*\|_2} = \frac{N_t}{D_t}$$

← Numerator

← Denominator

$$N_{t+1} = \vec{w}^{t+1} \cdot \vec{w}^* = \vec{w}^t \cdot \vec{w}^* + \vec{a}_{j_t} \cdot \vec{w}^*$$

$$\geq N_t + r_p \|\vec{w}^*\|_2$$

$$\frac{D_{t+1}^2}{\|\vec{w}^*\|_2^2} = \|\vec{w}^{t+1}\|_2^2 = (\vec{w}^t + \vec{a}_{j_t}) \cdot (\vec{w}^t + \vec{a}_{j_t})$$

$$= \|\vec{w}^t\|_2^2 + 2(\vec{w}^t \cdot \vec{a}_{j_t}) + \|a_{j_t}\|_2^2$$

$$\leq \frac{D_t^2}{\|w^*\|_2^2} + 1.$$

$$\therefore N_T \geq T \cdot r_p \|\vec{w}^*\|_2$$

$$D_T \leq \sqrt{T} \|\vec{w}^*\|_2$$

$$\therefore \Phi_T = \text{cosine}(\vec{w}^T, \vec{w}^*) \geq \frac{T \cdot r_p}{\sqrt{T}} = \sqrt{T}\, r_p$$

$$1 \geq \Phi_T \geq \sqrt{T}\, r_p \qquad T \leq \frac{1}{r_p^2}$$

$$T = O\left(\frac{1}{r_p^2}\right)$$

□

In this case Perceptron Algorithm does as well as the more sophisticated learning algorithm.

SUPPORT VECTOR MACHINE:

The two algorithms described earlier do not guarantee optimality of the margin computed.

Optimization Problem:

$$r_p = \max_{\vec{\omega}^*} \min_i \frac{\vec{\omega}^* \cdot \vec{a}_i}{\|\vec{\omega}^*\|_2 \, \|\vec{a}_i\|_2}$$

← Nonconvex ...

Since ~~$\forall i$~~ we can normalize the sample points such that

$$\|\vec{a}_i\|_2 = 1$$

$$r_p = \max_{\vec{\omega}^*} \min_i \frac{\vec{\omega}^*}{\|\vec{\omega}^*\|_2} \cdot a_i$$

Let $\vec{V} = \dfrac{\vec{\omega}}{\|\vec{\omega}\| \, r_p}$

An equivalent problem is

minimize $|\vec{V}|$ subject to $\underbrace{\vec{V} \cdot \vec{a}_i \geq 1 \quad \forall i}_{\text{Linear constraints}}$

↓ convex but not smooth

SVM - Large Margin Classifier Problem is

minimize $|\vec{V}|^2$

subject to $\vec{V} \cdot \vec{a}_i \geq 1 \quad \forall i$

LAGRANGE DUALITY.

SVM → Maximum Margin Classifiers

Constrained Optimization Problem.

$$\min_w f(w) \quad \text{subject to} \quad h_i(w) = 0 \quad \forall i \; i = 1, \dots \ell.$$

Lagrangian

$$\mathcal{L}(w, \beta) = f(w) + \sum_{i=1}^{\ell} \beta_i h_i(w)$$

↳ Lagrange Multipliers.

$$\frac{\partial \mathcal{L}}{\partial w_i} = 0 \quad \& \quad \frac{\partial \mathcal{L}}{\partial \beta_i} = 0 \quad \forall i \; i = 1, \dots \ell.$$

Generalized Lagrangian

$$\min_w f(w) \quad \text{s.t.} \quad g_i(w) \le 0 \quad \forall i = 1, \dots k$$
$$h_i(w) = 0 \quad \forall i = 1, \dots, \ell.$$

Generalized Lagrangian:

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{\ell} \beta_i h_i(w).$$

P: Primal →

$$\theta_p(w) = \max_{\alpha, \beta : \alpha_i \ge 0} \mathcal{L}(w, \alpha, \beta)$$

$\forall w$, $w \neq$ feasible

$g_i(w) > 0$

$h_i(w) \neq 0$

$$\theta_p(w) = \max_{\alpha, \beta : \alpha_i \ge 0} \mathcal{L}(w, \alpha, \beta)$$

$$= \max_{\alpha, \beta : \alpha_i \ge 0} f(w) + \sum_{i=1}^{k} \alpha_i g_i + \sum_{i=1}^{\ell} \beta_i h_i$$

$\to \infty$

$$\theta_p(\omega) = \begin{cases} f(\omega) & \text{if } \omega \text{ satisfies primal constraints} \\ & (\& \ \omega = \text{feasible}) \\ \infty & (\text{if } \omega = \text{infeasible}) \end{cases}$$

$$\min_{\omega} \theta_p(\omega) = \min_{\omega} \max_{\alpha,\beta:\alpha_i \geq 0} \mathcal{L}(\omega, \alpha, \beta)$$

$$= p^* \leftarrow \text{Value of the } \underline{\text{primal}} \text{ problem.}$$

DUALITY

$$\theta_{\mathcal{D}}(\omega) = \min_{\omega} \mathcal{L}(\omega, \alpha, \beta)$$

$$\max_{\alpha,\beta:\alpha_i \geq 0} \theta_{\mathcal{D}}(\omega) = \max_{\alpha,\beta:\alpha_i \geq 0} \min_{\omega} \mathcal{L}(\omega, \alpha, \beta)$$

$$= d^* \leftarrow \text{Value of the } \underline{\text{dual}} \text{ problem.}$$

$$\boxed{\max \min \leq \min \max}$$

$$d^* = \max_{\substack{\alpha,\beta \\ \alpha_i \geq 0}} \min_{\omega} \mathcal{L}(\omega, \alpha, \beta) \leq \min_{\omega} \max_{\substack{\alpha,\beta \\ \alpha_i \geq 0}} \mathcal{L}(\omega, \alpha, \beta)$$

$$= p^*$$

THEOREM (Karush-Kuhn-Tucker: KKT)

1) f and g_i = convex
2) h_i = affine
3) g_i's are strictly feasible
   $\exists_\omega \ g_i(\omega) < 0$ for all i

$\Rightarrow \quad d^* = p^* = \mathcal{L}(\omega^*, \alpha^*, \beta^*)$

ω^* = soln to the primal problem

α^*, β^* = soln to the dual problem

$\Leftrightarrow$ Satisfy KKT cond^{n}s

$$\begin{cases} \dfrac{\partial}{\partial \omega_i} \mathcal{L}(\omega, \alpha, \beta)\Big|_{\omega^*, \alpha^*, \beta^*} = 0 \quad \forall_i \\ \dfrac{\partial}{\partial \beta_i} \mathcal{L}(\omega, \alpha, \beta)\Big|_{\omega^*, \alpha^*, \beta^*} = 0 \quad \forall_i \end{cases} \quad \& \quad \begin{cases} \alpha_i^* g_i(\omega^*) = 0 \quad \forall_i \\ g_i(\omega^*) \leq 0 \quad \forall_i \\ \alpha^* \geq 0 \quad \forall_i \end{cases}$$

Optimal Margin Classifier:

$$\min \frac{1}{2}\|\vec{v}\|_2^2 \quad \text{s.t.} \quad \vec{v}\cdot\vec{a}_i \geq 1 \quad \forall_i$$

$$\mathcal{L}(\vec{v},\alpha) = \frac{1}{2}\vec{v}\cdot\vec{v} - \sum_i \alpha_i[\vec{v}\cdot\vec{a}_i - 1]$$

$$\nabla_v \mathcal{L} = \vec{v} - \sum_i \alpha_i \vec{a}_i = 0 \quad \Rightarrow \quad \vec{v} = \sum \alpha_i \vec{a}_i \;, \quad \alpha_i \geq 0$$

↳ Positive combination of $\vec{a}_i$'s

$$\mathcal{L}(\vec{v},\alpha) = \frac{1}{2}\left(\sum_i \alpha_i \vec{a}_i\right)\cdot\left(\sum_i \alpha_i \vec{a}_i\right) - \sum \alpha_i\left[\left(\sum_i \alpha_i \vec{a}_i\right)\cdot\vec{a}_i - 1\right]$$

$$= \frac{1}{2}\sum_{ij}\alpha_i\alpha_j(\vec{a}_i\cdot\vec{a}_j) - \sum_{i,j}\alpha_i\alpha_j(\vec{a}_i\cdot\vec{a}_j) + \sum_i \alpha_i$$

$$= \sum_i \alpha_i - \frac{1}{2}\sum_{ij}\alpha_i\alpha_j(\vec{a}_i\cdot\vec{a}_j)$$

DUAL

$$\Rightarrow \quad \max_{\alpha} \; W(\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_{ij}\alpha_i\alpha_j(\vec{a}_i\cdot\vec{a}_j)$$

subject to $\alpha_i \geq 0$

Soln $= \alpha^*$

$$\vec{v}^* = \sum \alpha_i^* \vec{a}_i = \frac{\vec{\omega}}{\|\vec{\omega}\|_2 r_p}$$

Recover $\vec{\omega} = (\vec{\theta}, -b) \longrightarrow \vec{\theta}\cdot\vec{x}_i - b = l_i$ □