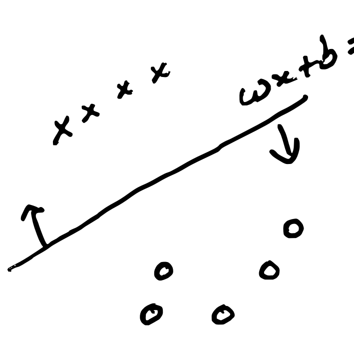
Perceptron algorithm

Ch5 Mohri et al textbook

$\to$ $w^{(t+1)} = w^{(t)} + x_i y_i$

$w^{(1)} = 0$

$x_i, w \in \mathbb{R}^d$



where

$$\underbrace{y_i (\langle w, x_i \rangle + b)}\ < 0$$

$$y_i (\langle w, x_i \rangle + b) \geq 1$$

Logic

$$y_i \langle w^{(t+1)}, x_i \rangle = y_i \langle w^{(t)}, x_i \rangle + \|x_i\|^2$$

Thm: Perceptron algorithm runs for no more than $R^2 B^2$ steps, where

$$R = \sup_x \|x\| \checkmark$$

$$B = \|w^*\|$$

Realizability

w^* is such that $\underline{y_i \langle \boxed{w^*}, x_i \rangle \geq 1 \quad \forall\, i \in [m]}$

$$B = \inf \{ \|w\| : \underline{y_i \langle w, x_i \rangle \geq 1 \ \forall i \in [m]} \}$$

Proof: T max # of iterations

$$w^{(t+1)} = w^{(t)} + y_i x_i$$

$$\|w^{(t+1)}\|^2 = \|w^{(t)} + y_i x_i\|^2$$

$$= \|w^{(t)}\|^2 + \|x_i\|^2 + \underline{2 y_i \langle w^{(t)}, x_i \rangle}$$

$$\|w^{(t+1)}\|^2 \leq \|w^{(t)}\|^2 + \|x_i\|^2$$

$$\leq \|w^{(t)}\|^2 + R^2$$

$$\underline{\|w^{(T+1)}\|^2} \leq \|w^{(T)}\|^2 + R^2$$

$$\leq \|w^{(T-1)}\|^2 + 2R^2$$

$$\vdots$$

$$\leq \underline{TR^2} \qquad (\because \|w^{(1)}\| = 0)$$

$$\langle w^{(t+1)}, w^* \rangle = \langle w^{(t)} + y_i x_i, w^* \rangle$$

$$= \langle w^{(t)}, w^* \rangle + \underline{y_i \langle w^*, x_i \rangle}$$

($\because$ realizability)

$$\geq \langle w^{(t)}, w^* \rangle + 1$$

$$\underbrace{\langle w^{(t+1)} - w^{(t)}, w^* \rangle} \geq 1$$

$$|\langle \underline{w^{(T+1)}}, \underline{w^*} \rangle| = \left| \sum_{t=1}^{T} \langle w^{(t+1)} - w^{(t)}, w^* \rangle \right|$$

$$(\because w^{(1)} = 0)$$

$$\geq T$$

Cauchy Schwarz

$$T \leq |\langle w^{(T+1)}, w^* \rangle| \leq \|w^{(T+1)}\| \|w^*\|$$

$$\leq \|w^{(T+1)}\| B$$

$$\leq \sqrt{T} R B$$

$$\underline{T \leq (RB)^2}$$

$\text{sgn}(\langle w, x \rangle + b)$

$\nexists\ (w, b)$ s.t.

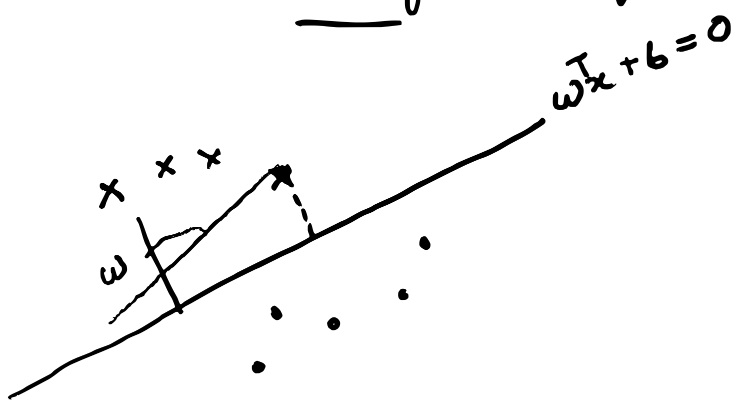
$$y_i (\langle w, x_i \rangle + b) \geq 0 \quad \forall\, i$$

$$w^{(t+1)} = w^{(t)} + y_i x_i$$

$$u^{(t+1)} = u^{(t)} + y_i' x_i'$$

$$\|w^{(t+1)} - u^{(t+1)}\| \leq 2R$$

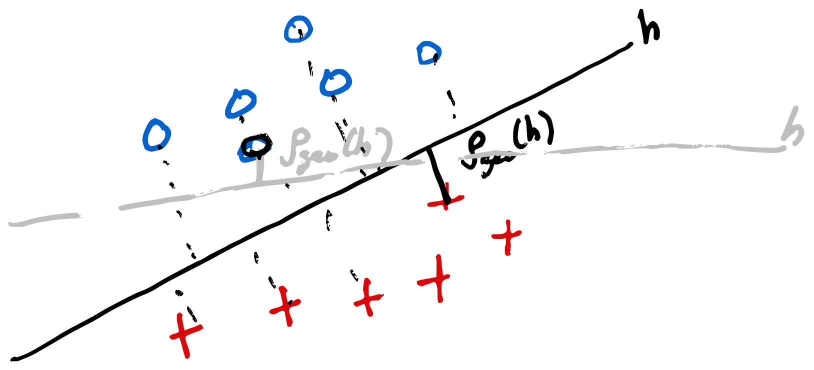
Maximum-margin classification



$$\rho_{geo}(\vec{h}) = \min_{i\in[m]} d_h(x_i)$$

$$d_h(x_i) = \frac{|w^T x_i + b|}{\|w\|} \quad ✓$$

Ex: proof



SVM

$$\max_{w,b} \min_{x_i\in S} \frac{|\langle w, x_i\rangle + b|}{\|w\|} \quad ✓$$

subject to

$$y_i(\langle w, x_i\rangle + b) \geq 0 \quad \forall i \in [m]$$

$$=$$

$$\max_{w,b} \min_{i\in[m]} \frac{y_i(\langle w, x_i\rangle + b)}{\|w\|} \qquad (C)$$

$$=$$

$$\max_{w,b} \frac{1}{\|w\|} \qquad (A)$$

subject to $\min_i y_i(\langle w, x_i\rangle + b) = 1$

$$=$$

$$\max_{w,b} \frac{1}{\|w\|} \qquad (B)$$

subject to $y_i(\langle w, x_i\rangle + b) \geq 1 \quad \forall i$

$$=$$

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

subject to $y_i(\langle w, x_i\rangle + b) \geq 1$

HARD SVM

→ Convex Ex.

Convex optimization

Lemma: if f is convex and differentiable,

$$f(x) - f(y) \geq \langle \nabla f(x), x-y \rangle \checkmark$$

iff condition for convexity

Review — Constrained optimization

Primal

$$\min_{w \in W} f(w)$$

subject to $g_i(w) \leq 0$

$i = 1, 2, \ldots, m$

Defn: Lagrangian

$$\mathcal{L}(w, \alpha) = f(w) + \sum_{i=1}^{m} \alpha_i g_i(w)$$

$$\alpha = [\alpha_1, \ldots, \alpha_m]^T$$

Defn: Dual function

$$F(\alpha) = \min_{w} \mathcal{L}(w, \alpha)$$

$$= f(w^*) + \sum_{i=1}^{m} \alpha_i g_i(w^*)$$

$$w^* = \underset{w}{\operatorname{argmin}} \mathcal{L}(w, \alpha)$$

Convex function: f is convex if

$$f(tw_1 + (1-t)w_2) \leq tf(w_1) + (1-t)f(w_2)$$

$$t \in (0, 1)$$

Dual problem

$$\max F(\alpha)$$

subject to $\alpha_i \geq 0$

$i = 1, 2, \ldots, m$

KKT

Karush Kuhn Tucker

Necessary and sufficient conditions for existence of unique solutions to convex optimization problems:

w^* is a minimizer of Primal problem, iff

$\rightarrow \exists\, w \in W \; g_i(w) \leq 0$ (Slater's condition)

$\rightarrow \nabla_w \mathcal{L}(w^*, \alpha^*) = 0$

$\rightarrow \nabla_\alpha \mathcal{L}(w^*, \alpha^*) = 0$

$\rightarrow \sum_{i=1}^{m} \alpha_i^* g_i(w^*) = 0$

$$\Rightarrow \alpha_i^* g_i(w^*) = 0 \quad \forall\, i \in [m]$$

Solution of soft/hard SVM

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + \cancel{\lambda \sum_{i=1}^{m} \xi_i}$$

subject to $\quad y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i$

$\xi_i \geq 0 \qquad i = 1, 2, \ldots, m$

KKT conditions

$$w = \sum_{i=1}^{m} \alpha_i y_i x_i$$

$\alpha_i = 0$ or $y_i(\langle w, x_i \rangle + b) = 1 - \xi_i$

$\beta_i = 0$ or $\xi_i = 0$

$\alpha_i + \beta_i = \lambda$

From dual form:

Solve for α