Primal problem:

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

Subject to

$$y_i(w_i \cdot x_i + b) \geq 1 - \boxed{\xi_i}$$

γ_i and $\xi_i \geq 0 \qquad i \in [m].$

$\xi = [\xi_1, \dots, \xi_m]^T$

$\xi_i = 0$ HARD SVM   Soft SVM

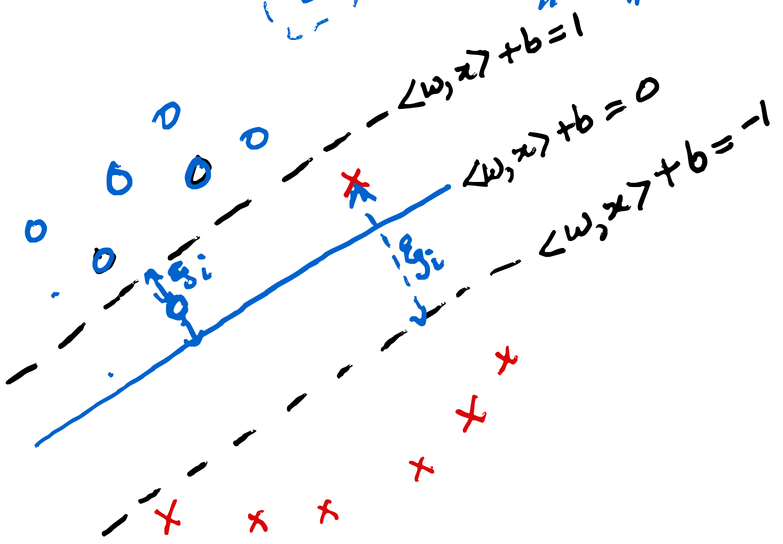
Today

1. SVM convex optimization
2. what are support vectors?
3. Analysis: generalization bounds for SVM

Material: Ch 5 of Mohri et al
SVM in Hastie, Tibshirani

Margin: $\frac{1}{\|w^*\|}$

$$\min_i \frac{|\langle w, x_i \rangle + b|}{\|w\|}$$



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

w^* is a minimizer of Primal problem, iff

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

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

Complementarity constraints

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

$$\Longrightarrow \quad \alpha_i^* g_i(w^*) = 0$$

Primal problem

$$\min_w f(w)$$

$$g_i(w) \leq 0 \qquad i \in [m]$$

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

$\mathcal{L}(w, b, \xi, \alpha, \beta)$

$$= \underline{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i} +$$

$$\sum_{i=1}^{m} \alpha_i \underbrace{\left(1-\xi_i - y_i(\langle w_i, x_i\rangle + b)\right)}$$

$$- \sum_{i=1}^{m} \beta_i \xi_i$$

(α_i, β_i), $i \in M$ Dual variables

Constraints:

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

✓ $-\xi_i \leq 0$

$\nabla_w \mathcal{L} = 0 \Rightarrow w = \sum_{i=1}^{m} \alpha_i y_i x_i$ ✓

$\nabla_b \mathcal{L} = 0$ $\sum_{i=1}^{m} \alpha_i y_i = 0$ ✓

$\nabla_{\xi_i} \mathcal{L} = 0$ $\alpha_i + \beta_i = C$

Complementarity:

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

$\alpha_i = 0$

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

$\forall i \in [m]$

x_i are called support vectors for any i when $\alpha_i \neq 0$

$$\mathcal{L} = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

$$\checkmark \ \underline{\omega = \sum_{i=1}^{m} \alpha_i \underline{y_i x_i}}$$

Dual <u>problem</u>

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

$$\underline{C} \geq \alpha_i \geq 0 \qquad \forall \ i \in [m]$$

<u>QP</u>

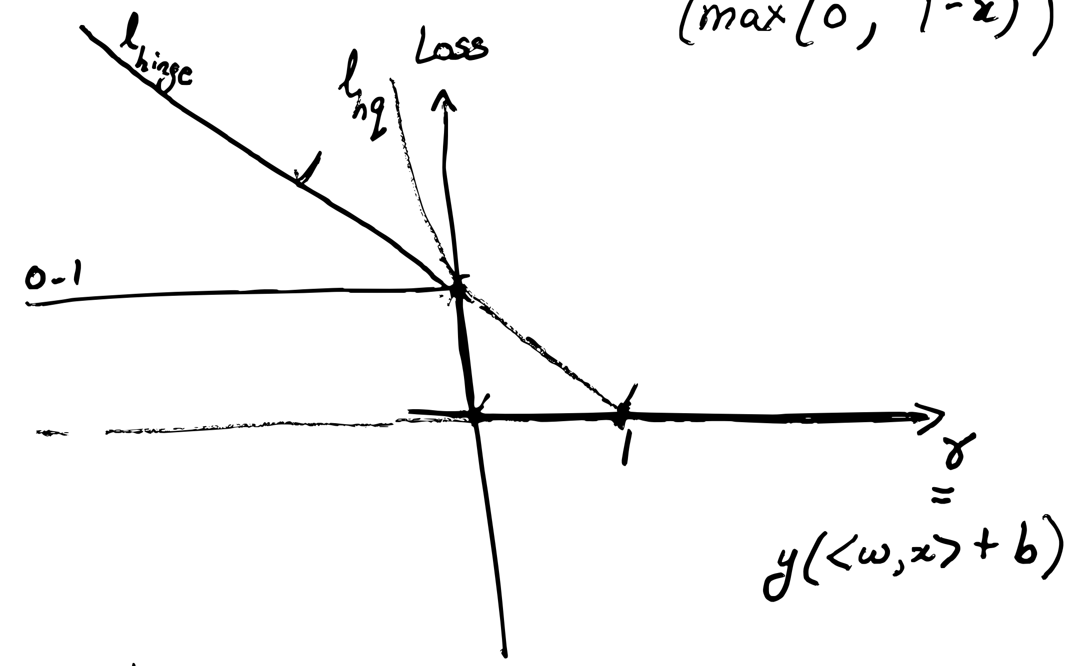
Margin loss function

$$\min\left(1,\ \max\left(0,\ 1-\frac{y(\langle w,x\rangle+b)}{\rho}\right)\right) \checkmark$$

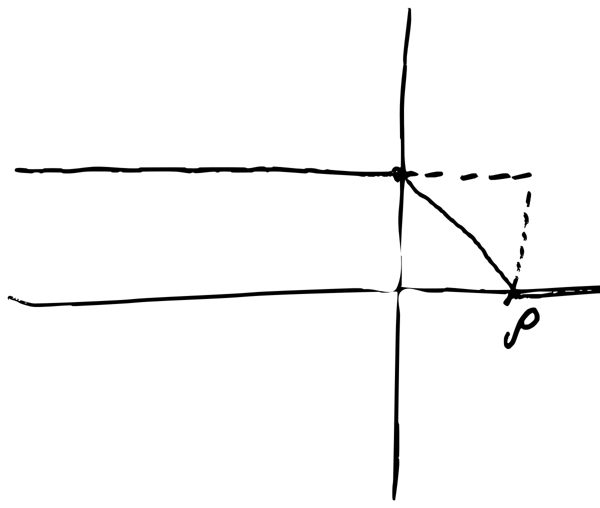
Hinge loss:

$$\ell_{hinge}(x) = \max(0, 1-x)$$

$$(\max(0, 1-x))^2$$



0-1 loss minimization is NP-hard



$$\rightarrow \hat{R}_{S,\rho}(h) = \sum_{i=1}^{m} \min\left(1, \max\left(0, \frac{(\langle w,x_i\rangle+b)y_i}{\rho}\right)\right)$$

(w,b)

$$\leq \underbrace{\sum_{i=1}^{m} \mathbb{1}_{\{(x_i,y_i)\in S:\ (\langle w,x_i\rangle+b)y_i \geq \rho\}}}_{\hat{R}_S(h)}$$

$$\mathop{\mathbb{E}}_{S\sim D^m} \hat{R}_S(h)$$

$$\parallel$$

$$R_S(h) \leq \hat{R}_{S,\rho}(h) + \_\_\_\_$$

$$\leq \hat{R}_S(h) + \_\_\_\_$$

If there is an "appropriate" margin ρ for the data distribution $\mathcal{D}$ then, SVM problem generalizes with a bound that does not "explicitly" depend on d.

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

Rademacher complexity

PAC

$$\underset{S}{\mathrm{Rad}}(\mathcal{H}) = \frac{1}{m} \underset{\sigma}{E}\left[\sup_{h \in \mathcal{H}} \sum_{i=1}^{m} \sigma_i h(x_i)\right]$$

$$\sigma = \{\sigma_1, \ldots, \sigma_m\}$$

$$\sigma_i = \begin{cases} 1 & \text{Probability } 1/2 \\ -1 & \text{Probality } 1/2 \end{cases}$$

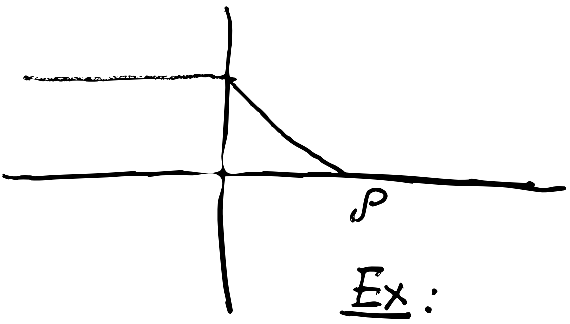
$\sum_{i=1}^{m} \sigma_i h(x_i)$: "correlation" between function output & noise

$$\mathcal{H} = \{ h(\cdot, w, b) : \langle w, x \rangle + b = h(x) \}$$

$$\mathcal{H} = \{ h(\cdot, w, b) : h(x) = \langle w, x \rangle + b, \ \|h(x)\| < \Lambda \}$$

Class of Classifiers is

$\tilde{\mathcal{H}} : \phi \circ \mathcal{H}$



e.g.

ϕ : sgn

Ex: $1/\rho$

ϕ : hinge loss

$$\tilde{\mathcal{H}} = \left\{ h(\cdot, w, b) : h(x) = \min\left(1, \max\left\{\frac{\langle w, x \rangle + b}{\rho}, 0\right\}\right) \right\}$$

Thm: If Φ is ℓ-lipschitz,

$$\mathrm{Rad}_S\left(\boxed{\Phi \circ \mathcal{H}}\right) \leq \ell \, \mathrm{Rad}_S(\mathcal{H})$$

Φ is ℓ-lipschitz if

$$|\Phi(x) - \Phi(y)| \leq \ell \, \|x - y\| \quad \forall x, y$$

$$\ell := \sup_x \|\nabla \Phi(x)\|$$

Next time: 1) $\mathrm{Rad}_S(\mathcal{H}) < C$

linear

2) Use thm