SVMs & Duality

General Constrained Optimization Problem:

$$\min_x \; f_0(x)$$

$$\text{st} \quad f_i(x) \le 0 \quad , i = 1,2,\dots m$$

$$\text{and} \quad h_i(x) = 0 \quad , i = 1,2,\dots p$$

① Primal problem

Here $x \in \mathbb{R}^n$ (i.e. $f_i : \mathbb{R}^n \to \mathbb{R}$, $h_i \in \mathbb{R}^n \to \mathbb{R}$)

Lagrangian: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$\underline{L}(\underline{x}, \underline{\lambda}, \underline{\nu}) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

where λ_i is Lagrange Multiplier for i-th inequality constraint

& ν_i " " " " i-th equality constraint

λ & ν are also called dual variables

Lagrange Dual Function

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

λ & ν are dual feasible if $\underline{\lambda} \ge 0$ & $g(\lambda, \nu) > -\infty$

<u>Fact 1</u> $\quad g(\lambda, \nu) \le p^*$ for any dual feasible λ, ν
(where p^* is primal optimal, i.e. optimal value of ①, $p^* = f_0(x^*)$)

<u>Fact 2</u> $\quad$ If $\exists$ dual feasible λ^*, ν^* ($\lambda^* \ge 0$) and primal feasible x^* st $g(\lambda^*, \nu^*) = p^* = f_0(x^*)$ then strong duality is said to hold

<u>Dual Problem</u>

$$\max_{\lambda, \nu} \; g(\lambda, \nu)$$

$$\text{st} \quad \lambda \ge 0 \longrightarrow (\lambda_i \ge 0, \; i = 1,2,\dots m)$$

Suppose strong duality holds for λ^*, ν^*, x^*

Then $f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x L(x, \lambda^*, \nu^*)$

$$= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right)$$

$$\leq f_0(x^*) + \underbrace{\sum_{i=1}^{m} \underbrace{\lambda_i^*}_{\geq 0} \underbrace{f_i(x^*)}_{\leq 0}}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* \underbrace{h_i(x^*)}_{=0}}_{=0}$$

$$\leq f_0(x^*)$$

So the above two inequalities must hold with equality

1. x^* must be minimizer of $L(x, \lambda^*, \nu^*)$
2. $\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0 \Rightarrow \lambda_i^* f_i(x^*) = 0$ for all $i = 1, 2, \ldots m$

   If $\lambda_i^* > 0$, then $f_i(x^*) = 0$

   If $f_i(x^*) < 0$, then $\lambda_i^* = 0$

Karush Kuhn Tucker (KKT) conditions: Provide certificate of optimality when problem (1) is convex

KKT Conditions for x^* and (λ^*, ν^*):

x^* Primal feasible: $\begin{cases} f_i(x^*) \leq 0, & i = 1, 2, \ldots m \\ h_i(x^*) = 0, & i = 1, 2, \ldots p \end{cases}$

λ^* Dual feasible: $\lambda_i^* \geq 0$, $i = 1, 2, \ldots m$

Complementary Slackness: $\lambda_i^* f_i(x^*) = 0$, $i = 1, 2, \ldots m$

$\rightarrow x^* = \operatorname{argmin} L(x, \lambda^*, \nu^*)$

$$\rightarrow \nabla_x f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_x f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla_x h_i(x^*) = 0$$

SVM Problem

$$\min_{w, w_0} \frac{1}{2} \|w\|^2$$

st $y_i(w^T x_i + w_0) \geq 1$, for $i = 1, 2, \ldots N$

$$\underbrace{\boxed{1 - y_i(w^T x_i + w_0)}}_{f_i(w)} \leq 0 \quad \text{for } i = 1, 2 \ldots N$$

Lagrangian

$$L(w, w_0, \alpha) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^{N} \alpha_i \left(1 - y_i(w^T x_i + w_0)\right)$$

$$\nabla_w L(w, w_0, \alpha) = w + \sum_{i=1}^{N} -\alpha_i y_i x_i = 0 \Rightarrow \boxed{w = \sum_{i=1}^{N} \alpha_i y_i x_i}$$

$$\nabla_{w_0} L(w, w_0, \alpha) = \sum_{i=1}^{N} -\alpha_i y_i = 0 \Rightarrow \boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$$

Dual function $g(\alpha) = \inf_{w, w_0} L(w, w_0, \alpha)$

$$g(\alpha) = \frac{1}{2} \left\| \sum_{i=1}^{n} \alpha_i y_i x_i \right\|^2 + \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i y_i x_i^T \left(\sum_{j=1}^{N} \alpha_j y_j x_j \right) + \underbrace{\sum_{i=1}^{N} \cancel{\alpha_i y_i} w_0}_{0}$$

$$g(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

SVM Dual:

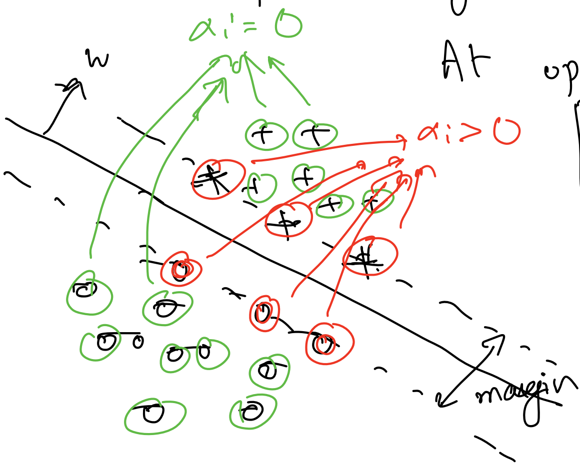
$$\boxed{\begin{aligned} &\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ &\text{such that} \quad \alpha_i \geq 0, \quad i = 1, 2 \ldots N \\ &\qquad \& \sum_{i=1}^{N} \alpha_i y_i = 0 \end{aligned}}$$

Note that:

$$\frac{1}{2} \left\| \sum_{i=1}^{N} \alpha_i y_i x_i \right\|^2 = \frac{1}{2} \left(\sum_{i=1}^{N} \alpha_i y_i x_i^T \right) \left(\sum_{j=1}^{N} \alpha_j y_j x_j \right) \qquad (\|z\|_2^2 = z^T z)$$

$$= \frac{1}{2} \sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Complementary Slackness



At optimality, α_i, w, w_0 must satisfy

$$\boxed{\alpha_i \left(1 - y_i (w^T x_i + w_0)\right) = 0 \quad \text{for all } i}$$

If $\alpha_i > 0$ then $\boxed{y_i (w^T x_i + w_0) = 1}$

If $y_i (w^T x_i + w_0) > 1$, then $\alpha_i = 0$

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

w is linear combination of "support" vectors

$\alpha_i > 0$

Recall the original separable SVM formulation:

$$\max_{w, w_0} \; C \quad \text{st} \quad y_i (w^T x_i + w_0) \geq C \|w\|$$

In order to allow non-separable data sets, relax above inequality to have a slack, i.e.,

$$y_i (w^T x_i + w_0) \geq C (1 - \xi_i) \|w\| , \quad \xi_i \geq 0$$

Same as earlier, we can fix $C \|w\| = 1$

Non-linearly separable SVM:

Primal: $\displaystyle \min_{w, w_0, \xi} \; \frac{1}{2} \|w\|^2$

$$y_i (w^T x_i + w_0) \geq 1 - \xi_i \quad , i = 1, 2, \ldots N$$

$$\xi_i \geq 0$$

$$\sum_{i=1}^{N} \xi_i \leq \text{constant}$$

(Non-linearly separable) SVM Primal:

$$\boxed{\begin{aligned} &\min_{w, w_0, \xi} \ \frac{1}{2}\|w\|^2 + r\left(\sum_{i=1}^{N} \xi_i\right) \\ &1 - \xi_i - y_i(w^T x_i + w_0) \leq 0, \quad i = 1, 2, \ldots N \\ &\xi_i \geq 0, \quad i = 1, 2, \ldots N \end{aligned}}$$

(Non-linearly separable) SVM Dual:

$$\boxed{\begin{aligned} &\max_{\alpha} \ \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j x_i^T x_j \\ &0 \leq \alpha_i \leq r, \quad i = 1, 2, \ldots N \\ &\sum_{i=1}^{N} \alpha_i y_i = 0 \end{aligned}}$$

Exercise: Derive the above SVM Dual