Machine Learning - SVM
The Primal and Dual Problems[1]

Jaesik Choi

Ulsan National Institute of Science and Technology

jaesik@unist.ac.kr

[1] These slides are based on the lecture notes of Andrew Ng

The Primal optimization problem - Lagrangian

The primal optimization problem:

$$\begin{aligned} \min_{w} \quad & f(w) \\ s.t. \quad & g_i(w) \leq 0, i = 1, \cdots, k. \end{aligned}$$

The generalized Lagrangian:

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

where α_i's is the Lagrange multipliers.

The Primal optimization problem

Let's optimize the $\mathcal{L}$ by α_i's first:

$$\theta_{\mathcal{P}}(w) = \max_{\alpha, \alpha_i \geq 0} \mathcal{L}(w, \alpha).$$

When w violates any of the primal constraints (e.g., $g_i(w) > 0$ for some i), then you should be able to verify that:

$$\theta_{\mathcal{P}}(w) = \infty.$$

Thus, $\theta_{\mathcal{P}}(w) = f(w)$, if w satisfies primal constraints. Thus, the primal optimization problem becomes:

$$\min_w \theta_{\mathcal{P}}(w) = \min_w \max_{\alpha, \alpha_i \geq 0} \mathcal{L}(w, \alpha).$$

The Dual optimization problem

Let's optimize the $\mathcal{L}$ by w's first:

$$\theta_{\mathcal{D}}(\alpha) = \min_{w} \mathcal{L}(w, \alpha).$$

The dual optimization problem becomes:

$$\max_{\alpha, \alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha) = \max_{\alpha, \alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha).$$

The Primal vs The Dual

In general, the primal is greater than equal to the dual:

$$\min_{w} \max_{\alpha, \alpha_i \geq 0} \mathcal{L}(w, \alpha) \geq \max_{\alpha, \alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha)$$

The equality holds when Karush-Kuhn-Tucker (KKT), conditions hold for some w^*, α,

$$\begin{aligned}
\frac{\partial}{\partial w_i} \mathcal{L}(w^*, \alpha*) &= 0, i = 1, \cdots, n \\
\alpha_i \cdot g_i(w^*) &= 0, i = 1, \cdots, k \\
g_i(w^*) &\leq 0, i = 1, \cdots, k \\
\alpha_i &\geq 0, i = 1, \cdots, k.
\end{aligned}$$

The Primal Dual in SVM

$$\mathcal{L}(w, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i (y^i w x^i - 1).$$

$$\mathcal{L}(w, \alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j}^{m} y^i y^j \alpha_i \alpha_j (x_i)^T x^j.$$

The End