Support Vector Machine

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Primal and Dual Problem

Primal Problem	Lagrange Dual Problem
$min_x f(x)$ $s.t.\ g(x) \le 0, h(x) = 0$	$max_{\alpha>0,\beta} d(\alpha,\beta)$ $s.t. \alpha > 0$
↓	↓
$min_x max_{\alpha\ge0,\beta} L(x,\alpha,\beta)$	$max_{\alpha\ge0,\beta} min_x L(x,\alpha,\beta)$

- Weak duality theorem
  - $d(\alpha,\beta) \le f(x^*) \quad for \quad \forall\alpha, \forall\beta$
  - $d^* = max_{\alpha\ge0,\beta} min_x L(x,\alpha,\beta) \le min_x max_{\alpha\ge0,\beta} L(x,\alpha,\beta) = p^*$
    - Maximizing the dual function provides the lower bound of $f(x^*)$
  - Duality gap = $f(x^*) - d(\alpha^*,\beta^*)$
- Strong duality
  - $d^* = max_{\alpha\ge0,\beta} min_x L(x,\alpha,\beta) = min_x max_{\alpha\ge0,\beta} L(x,\alpha,\beta) = p^*$
  - When Karush-Kunh-Tucker (KKT) Conditions are satisfied

KKT Condition and Strong Duality

- Strong duality
  - $d^* = max_{\alpha \geq 0, \beta} min_x L(x, \alpha, \beta) = min_x max_{\alpha \geq 0, \beta} L(x, \alpha, \beta) = p^*$
- Holds when KKT conditions are met
  - $\nabla L(x^*, \alpha^*, \beta^*) = 0$
  - $\alpha^* \geq 0$
  - $g(x^*) \leq 0$
  - $h(x^*) = 0$
  - $\alpha^* g(x^*) = 0$

Active Constraint
$\alpha^* = 0 \Rightarrow g(x^*) = 0$
Inactive Constraint
$g(x^*) < 0 \Rightarrow \alpha^* = 0$
→ Complementary Slackness

Primal Problem

$$min_x f(x)$$
$$s.t.\ g(x) \leq 0,$$
$$h(x) = 0$$

Strong Duality
$d^* = p^*$

Always

KKT Condition

For convex optimization

Primal and dual problems are equivalent for the constrained convex optimization

Dual Problem of SVM

Primal Problem

$$\min_x f(x)$$
$$s.t.\ g(x) \le 0,$$
$$h(x) = 0$$

Lagrange Prime Function

$$L(x, \alpha, \beta) = f(x) + \alpha g(x) + \beta h(x)$$

Primal Problem of Linearly Separable SVM

$$\min_{w,b} ||w||$$
$$s.t. (wx_j + b)y_j \ge 1, \forall j$$

$$\min_{w,b} \max_{\alpha \ge 0, \beta} \frac{1}{2} w \cdot w - \sum_j \alpha_j [(wx_j + b)y_j - 1]$$
$$s.t. \alpha_j \ge 0, for\ \forall j$$

- Linearly separable case
- Lagrange Prime Function
  - $L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_j \alpha_j [(wx_j + b)y_j - 1]$
- Lagrange Multiplier
  - $\alpha_j \ge 0, for\ \forall j$

Dual Problem of Linearly Separable SVM

$$\max_{\alpha \ge 0} \min_{w,b} \frac{1}{2} w \cdot w - \sum_j \alpha_j [(wx_j + b)y_j - 1]$$
$$s.t. \alpha_j \ge 0, for\ \forall j$$

KKT Condition to Eliminate the Duality Gap

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0, \frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0$$
$$\alpha_i \ge 0, \forall i$$
$$\alpha_i \left((wx_j + b)y_j - 1 \right) = 0, \forall i$$

Dual Representation of SVM

- $L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_j \alpha_j [(wx_j + b)y_j - 1]$
    - $= \frac{1}{2} ww \; - \sum_j \alpha_j y_j w x_j - b \sum_j \alpha_j y_j + \sum_j \alpha_j$
    - $= \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j - \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j - b \times 0 + \sum_j \alpha_j$
    - $= \sum_j \alpha_j - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$
- Again, a quadratic programming
- Once, α_j is known
    - $\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}_i$
    - $\alpha_i \left((wx_j + b)y_j - 1 \right) = 0$
- Now, we can find out the $\boldsymbol{w}$ and b again.
    - Why is this better?
    - Most of α_j are.....
    - Location of x is....
- Let's find out from the implementation…

Dual Problem of Linearly Separable SVM

$$max_{\alpha \geq 0} min_{w,b} \frac{1}{2} w \cdot w - \sum_j \alpha_j [(wx_j + b)y_j - 1]$$

$$s.t. \alpha_j \geq 0, for \ \forall j$$

KKT Condition to Eliminate the Duality Gap

$$\frac{\partial L(\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{\alpha})}{\partial \boldsymbol{w}} = 0 \, , \frac{\partial L(\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{\alpha})}{\partial b} = 0$$

$$\alpha_i \geq 0, \forall i$$

$$\alpha_i \left((wx_j + b)y_j - 1 \right) = 0, \forall i$$

$$\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}_i \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$