10-801: Advanced Topics in Graphical Models 10-801, Spring 2007

Optimal Margin Principle and Lagrangian Duality

Lecturer: Eric P. Xing *Scribes: Guang Xiang*

1 Lagrangian Duality

1.1 Primal and Dual Problems

The optimization problem (primal form) in the standard form is

$$\begin{gathered}\text{minimize } f_0(x), x \in \mathcal{D} \subseteq R^n \\ \text{s.t. } f_i(x) \leq 0,\ i \in \{1, 2, \cdots, m\} \\ g_i(x) = 0,\ i \in \{1, 2, \cdots, p\}\end{gathered}$$

The Lagrangian is defined for the optimization problem, which has the following form

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

where λ_i, ν_i are the Lagrange multipliers.

The Lagrange dual function can be defined over the Lagrangian as

$$g(\lambda, \nu) \quad = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$$

g is a concave function regardless of the convexity of the original optimization problem.

The optimal solutions of the primal and dual problems are

$$\begin{aligned} p^* &= \min_x \max_{\lambda,\nu} L(x, \lambda, \nu) \quad \text{(primal)} \\ d^* &= \max_{\lambda,\nu} \min_x L(x, \lambda, \nu) \quad \text{(dual)} \end{aligned}$$

1.2 Weak and Strong Duality

Weak Duality: The optimal value of the Lagrange dual problem d^* is the best lower bound on the optimal value of the primal problem p^*, i.e., $d^* \leq p^*$.

Strong Duality: the optimal values of the primal problem and dual problem agrees, i.e. $d^* = p^*$. Strong duality holds when

1. The primal problem is convex, or
2. Slater's condition holds: $\exists x \in$ **relint** $\mathcal{D}$, $f_i(x) < 0, i \in \{1, 2, \cdots, m\}$, $Ax = b$ (equality constraints)

1.3 Karush-Kuhn-Tucker (KKT) conditions

For nonconvex problems: for an optimization problem with differentiable objective and constraint functions, if strong duality holds, the optimal primal and dual values satisfy

$$\begin{aligned} f_i(x) &\leq 0, \quad i \in \{1, 2, \cdots, m\} \\ g_i(x) &= 0, \quad i \in \{1, 2, \cdots, p\} \\ \lambda_i &\geq 0, \quad i \in \{1, 2, \cdots, m\} \\ \lambda_i f_i(x) &= 0, \quad i \in \{1, 2, \cdots, m\} \\ \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla g_i(x) &= 0, \end{aligned}$$

For convex problems: for an optimization problem with differentiable objective and constraint functions, if the primal problem is convex and x, λ, ν satisfy the KKT conditions, then x and (λ, ν) are primal and dual optimal with the property of $p^* = d^*$.

2 Support Vector Machines

For the linearly separable case, the SVM aims at finding a hyperplane that maximizes the margin between the two opposite classes, which is equivalent to the following optimization problem

$$\begin{gathered} \min \tfrac{1}{2}\|w\|^2 \\ \text{s.t. } y_i(x_i \cdot w + b) - 1 \geq 0 \ \forall \ i \end{gathered}$$

The Lagrangian in primal form for this problem is

$$L_P = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{l} \alpha_i$$

Assuming the gradient of L_P with respect to w and b vanish, we have the following intermediate results

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

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$

Plugging these two equations into the primal problem gives us the dual form

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j$$

The solution to the separating hyperplane is found by minimizing L_P or maximizing L_D.

The KKT conditions for the primal problem are

$$\begin{aligned}
\tfrac{\partial L_P}{\partial w_v} = w_v - \textstyle\sum_i \alpha_i y_i x_{iv} &= 0 \quad v \in \{1, \cdots, d\} \\
\tfrac{\partial L_P}{\partial b} = -\textstyle\sum_i \alpha_i y_i &= 0 \\
y_i(x_i \cdot w + b) - 1 &\geq 0 \quad i \in \{1, \cdots, l\} \\
\alpha_i &\geq 0 \quad i \in \{1, \cdots, l\} \\
\alpha_i(y_i(x_i \cdot w + b) - 1) &= 0 \quad i \in \{1, \cdots, l\}
\end{aligned}$$

For the non-linearly separable case, we introduce a set of slack variables $\xi_i, i \in \{1, \cdots, l\}$ and define the constraint functions in a similar manner.

$$y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0,\ \xi_i \geq 0\ \forall i$$

The primal and dual problems in this case are

$$L_P = \frac{1}{2}\|w\|^2 + C\sum_i \xi_i - \sum_{i=1} \alpha_i\{y_i(x_i \cdot w + b) - 1 + \xi_i\} - \sum_{i=1} \mu_i \xi_i$$

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j$$

The KKT conditions for the primal problem under the non-separable scenario are

$$\begin{aligned}
\tfrac{\partial L_P}{\partial w_v} = w_v - \textstyle\sum_i \alpha_i y_i x_{iv} &= 0 \\
\tfrac{\partial L_P}{\partial b} = -\textstyle\sum_i \alpha_i y_i &= 0 \\
\tfrac{\partial L_P}{\partial \xi_i} = C - \alpha_i - \mu_i &= 0 \\
y_i(x_i \cdot w + b) - 1 + \xi_i &\geq 0 \\
\xi_i &\geq 0 \\
\alpha_i &\geq 0 \\
\mu_i &\geq 0 \\
\alpha_i\{y_i(x_i \cdot w + b) - 1 + \xi_i\} &= 0 \\
\mu_i \xi_i &= 0
\end{aligned}$$