Dual of a Linear Support Vector Machine Quadratic Program

C. Phillip Cloud

March 10, 2012

This short description borrows heavily from Professor Andrew Ng's lecture notes on support vector machines.

1 Lagrangians

Given the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}) && (1) \\ \text{s.t.} \quad & h_i(\mathbf{w}) = 0,\ i = 1, \ldots, n, && (2) \end{aligned}$$

that is, minimize $f(\mathbf{w})$ with respect to $\mathbf{w}$ such that some function $h_i(\mathbf{w})$ is equal to zero.

Define the Lagrangian to be

$$\mathcal{L}(\mathbf{w}, \boldsymbol{\beta}) = f(\mathbf{w}) + \sum_{i=1}^{n} \beta_i h_i(\mathbf{w}). \tag{3}$$

1.1 Primal Problem

Given a more general set of constraints,

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}) && (4) \\ \text{s.t.} \quad & g_i(\mathbf{w}) \leq 0,\ i = 1, \ldots, k && (5) \\ & h_i(\mathbf{w}) = 0,\ i = 1, \ldots, n && (6) \end{aligned}$$

we can define a more general form of the Lagrangian. We have

$$\mathcal{L}\left(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}\right) = f\left(\mathbf{w}\right) + \sum_{i=1}^{k} \alpha_i g_i\left(\mathbf{w}\right) + \sum_{i=1}^{n} \beta_i h_i\left(\mathbf{w}\right), \tag{7}$$

and we want to maximize this function with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$,

$$\boldsymbol{\theta}\left(\mathbf{w}\right) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} \mathcal{L}\left(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}\right). \tag{8}$$

If we want to minimize $\mathcal{L}$ with respect to $\mathbf{w}$ then we have

$$\min_{\mathbf{w}} \boldsymbol{\theta}\left(\mathbf{w}\right) = \min_{\mathbf{w}} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} f\left(\mathbf{w}\right) + \sum_{i=1}^{k} \alpha_i g_i\left(\mathbf{w}\right) + \sum_{i=1}^{n} \beta_i h_i\left(\mathbf{w}\right). \tag{9}$$

This is the value of the *primal* problem, which we will call

$$p^* = \min_{\mathbf{w}} \boldsymbol{\theta}\left(\mathbf{w}\right). \tag{10}$$

1.2 Dual Problem

Now define the dual of p^*, call it d^*, as

$$d^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} p^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} \min_{\mathbf{w}} \boldsymbol{\theta}\left(\mathbf{w}\right). \tag{11}$$

In the case that $p^* = d^*$, we have *strong duality*.

2 Support Vector Machines

For a linear *support vector machine* we have the following constrained optimization problem,

$$\min_{\boldsymbol{\alpha},\mathbf{w},b} \quad \tfrac{1}{2}\|\mathbf{w}\|_2^2 \tag{12}$$

$$\text{s.t.} \quad y_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) \geq 1, \; i = 1, \ldots, m. \tag{13}$$

This problem can also be seen as a quadratic program if we write equation 12 as

$$\min_{\boldsymbol{\alpha},\mathbf{w},b} \quad \tfrac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{I}_m\mathbf{w} + \mathbf{0}_m^{\mathrm{T}}\mathbf{w} \tag{14}$$

$$\text{s.t.} \quad -\mathbf{y} * \left(\mathbf{X}\mathbf{w} + b \cdot \mathbf{1}_m\right) + 1 \leq 0 \tag{15}$$

where $\mathbf{I}_m$ is the $m \times m$ identity matrix, $\mathbf{0}_m$ is the zero vector with m elements, $\mathbf{1}_m$ is the vector with elements all equal to 1, and $*$ is the Hadamard (element-wise) product of two vectors.

The constraint in equation 13 can be rewritten as

$$g_i(\mathbf{w}) = -y_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) + 1 \leq 0. \tag{16}$$

Plugging in equations 12 and 16 to the definition of the Lagrangian in equation 3, the primal of this problem is

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{i=1}^{m}\alpha_i\left(y_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) - 1\right). \tag{17}$$

The dual is found by taking the derivatives of $\mathcal{L}$ with respect to $\mathbf{w}$ and b and setting them equal to zero and then plugging the optimal $\mathbf{w}$ back in to the dual. For $\mathbf{w}$, we have

$$\begin{aligned}
\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) &= \mathbf{w} - \sum_{i=1}^{m}\alpha_i y_i \mathbf{x}_i = 0 && (18)\\
\mathbf{w} &= \sum_{i=1}^{m}\alpha_i y_i \mathbf{x}_i, && (19)
\end{aligned}$$

and for b,

$$\frac{\partial}{\partial b}\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \sum_{i=1}^{m}\alpha_i y_i = 0 \tag{20}$$

Plugging in the optimal $\mathbf{w}$ to equation 17 we see that the dual of a linear SVM quadratic program is

$$\begin{aligned}
\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{i=1}^{m}\left(\alpha_i y_i \mathbf{w}^{\mathrm{T}}\mathbf{x}_i + \alpha_i y_i b - \alpha_i\right) && (21)\\
&= \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{i=1}^{m}\alpha_i y_i \mathbf{w}^{\mathrm{T}}\mathbf{x}_i - b\cancel{\sum_{i=1}^{m}\alpha_i y_i} + \sum_{i=1}^{m}\alpha_i && (22)\\
&= \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{i=1}^{m}\alpha_i y_i \mathbf{w}^{\mathrm{T}}\mathbf{x}_i + \sum_{i=1}^{m}\alpha_i && (23)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\|\mathbf{w}\|_2^2 - \mathbf{w}^{\mathrm{T}} \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i + \sum_{i=1}^{m} \alpha_i && (24) \\
&= \frac{1}{2}\|\mathbf{w}\|_2^2 - \|\mathbf{w}\|_2^2 + \sum_{i=1}^{m} \alpha_i && (25) \\
&= \|\mathbf{w}\|_2^2 \left(\frac{1}{2} - 1\right) + \sum_{i=1}^{m} \alpha_i && (26) \\
&= \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\|\mathbf{w}\|_2^2 && (27) \\
&= \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j. && (28)
\end{aligned}
$$

Equation 28 shows that the dual depends only on the inner product of the vectors $\mathbf{x}_i$ and $\mathbf{x}_j$.