Machine Learning
CMPT 726

Mo Chen
SFU School of Computing Science
2025-10-31

Lagrangian Duality

Consider a generic constrained optimization problem:

$$\min_{\vec{\theta}} f\left(\vec{\theta}\right)$$

Objective function

subject to

$$g_i\left(\vec{\theta}\right) \leq 0 \ \forall i \in \{1, \dots, k\}$$

Inequality constraints

$$h_i\left(\vec{\theta}\right) = 0 \ \forall i \in \{1, \dots, l\}$$

Equality constraints

$S := \{\vec{\theta} | g_i(\vec{\theta}) \leq 0 \ \forall i \in \{1, \dots, k\} \text{ and } h_i(\vec{\theta}) = 0 \ \forall i \in \{1, \dots, l\}\}$ is known as the **feasible region**. A solution $\vec{\theta}$ is the feasible region is known as a **feasible solution**.

Generalized Lagrangian

The generalized Lagrangian turns a constrained optimization problem to an unconstrained optimization problem.

$$\mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right) = f\left(\vec{\theta}\right) + \sum_{i=1}^{k} \lambda_i g_i\left(\vec{\theta}\right) + \sum_{i=1}^{l} v_i h_i\left(\vec{\theta}\right)$$

Consider the function $\Phi\left(\vec{\theta}\right) = \max\limits_{\vec{\lambda},\vec{v}:\lambda_i \geq 0 \ \forall i} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$

Observe that:

$$\Phi\left(\vec{\theta}\right) = \begin{cases} f\left(\vec{\theta}\right), & g_i\left(\vec{\theta}\right) \leq 0 \ \forall i \in \{1, \dots, k\} \quad \text{and} \quad h_i\left(\vec{\theta}\right) = 0 \ \forall i \in \{1, \dots, l\} \\ \infty, & \text{otherwise} \end{cases}$$

Generalized Lagrangian

$$p^* = \min_{\vec{\theta}} \Phi\left(\vec{\theta}\right) = \min_{\vec{\theta}} \max_{\vec{\lambda},\vec{v}:\lambda_i \geq 0 \ \forall i} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$$

Since $\Phi(\vec{\theta}) = f(\vec{\theta})$ in the feasible region and is ∞ otherwise, the above is equivalent to the original constrained optimization problem.

The above is known as the **primal problem**.

Now consider the following optimization problem, which is different in that the min and max are swapped:

$$d^* = \max_{\vec{\lambda},\vec{v}:\lambda_i \geq 0 \ \forall i} \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$$

The above is known as the **dual problem**.

Weak Duality

Weak duality relates $p^* = \min_{\vec{\theta}} \max_{\vec{\lambda}, \vec{v}: \lambda_i \geq 0 \ \forall i} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$ to $d^* = \max_{\vec{\lambda}, \vec{v}: \lambda_i \geq 0 \ \forall i} \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$.

$\forall \vec{\theta}', \mathcal{L}(\vec{\theta}', \vec{\lambda}, \vec{v}) \geq \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$ at **every point** in the space of $\vec{\lambda}$ and $\vec{v}$. (1)

Consider a point $(\vec{\lambda}_0, \vec{v}_0)$ that maximizes the function $\min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$.

So, from (1), in particular, $\forall \vec{\theta}', \mathcal{L}(\vec{\theta}', \vec{\lambda}_0, \vec{v}_0) \geq \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}_0, \vec{v}_0\right) = \max_{\vec{\lambda}, \vec{v}} \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$. (2)

On the other hand, $\forall \vec{\theta}', \max_{\vec{\lambda}, \vec{v}} \mathcal{L}(\vec{\theta}', \vec{\lambda}, \vec{v}) \geq \mathcal{L}\left(\vec{\theta}', \vec{\lambda}_0, \vec{v}_0\right)$. (3)

Combining (2) and (3), we have $\forall \vec{\theta}', \max_{\vec{\lambda}, \vec{v}} \mathcal{L}(\vec{\theta}', \vec{\lambda}, \vec{v}) \geq \mathcal{L}\left(\vec{\theta}', \vec{\lambda}_0, \vec{v}_0\right) \geq \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}_0, \vec{v}_0\right) = \max_{\vec{\lambda}, \vec{v}} \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$. (4)

Weak Duality

From the previous slide, we have: $\forall \vec{\theta}', \max_{\vec{\lambda}, \vec{v}} \mathcal{L}(\vec{\theta}', \vec{\lambda}, \vec{v}) \geq \max_{\vec{\lambda}, \vec{v}} \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$. (4)

Consider a point $\vec{\theta}'_0$ that minimizes the function $\max_{\vec{\lambda}, \vec{v}} \mathcal{L}(\vec{\theta}', \vec{\lambda}, \vec{v})$.

From (4), in particular, $\max_{\vec{\lambda}, \vec{v}} \mathcal{L}(\vec{\theta}'_0, \vec{\lambda}, \vec{v}) \geq \max_{\vec{\lambda}, \vec{v}} \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$.

Therefore, $p^* = \min_{\vec{\theta}} \max_{\vec{\lambda}, \vec{v}} \mathcal{L}(\vec{\theta}, \vec{\lambda}, \vec{v}) \geq \max_{\vec{\lambda}, \vec{v}} \min_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right) = d^*$

Strong Duality

The quantity $p^* - d^*$ is known as the **duality gap**.

Under some conditions, the duality gap is zero, i.e. $p^* = d^*$.

In this case, we say **strong duality** holds.

Strong duality is nice because if it holds, solving the primal problem of $\min\limits_{\vec{\theta}} \max\limits_{\vec{\lambda},\vec{v}} \mathcal{L}(\vec{\theta}, \vec{\lambda}, \vec{v})$ is equivalent

to solving the dual problem of $\max\limits_{\vec{\lambda},\vec{v}} \min\limits_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$.

Sometimes the dual problem can be solved more efficiently than the primal problem.

Sometimes it reveals relationships between the optimal solution and the quantities given in the problem that are not obvious from the primal problem.

In the case of SVM, we will see both when we look at the dual problem. It also results in a natural way to generalize an SVM.

Dual Form of the SVM

Recall the primal form of the SVM: $\min\limits_{\vec{w},b} \frac{1}{2}\|\vec{w}\|_2^2$ s.t. $y_i(\vec{w}^\top \vec{x}_i - b) \geq 1\ \forall i$

Recall: For $\min\limits_{\vec{\theta}} f(\vec{\theta})$ s.t. $g_i(\vec{\theta}) \leq 0\ \forall i \in \{1, \dots, k\}$ and $h_i(\vec{\theta}) = 0\ \forall i \in \{1, \dots, l\}$, the dual problem is

$\max\limits_{\vec{\lambda},\vec{v}:\lambda_i \geq 0\ \forall i} \min\limits_{\vec{\theta}} \mathcal{L}(\vec{\theta}, \vec{\lambda}, \vec{v})$, where $\mathcal{L}(\vec{\theta}, \vec{\lambda}, \vec{v}) = f(\vec{\theta}) + \sum_{i=1}^{k} \lambda_i g_i(\vec{\theta}) + \sum_{i=1}^{l} v_i h_i(\vec{\theta})$

In the case of the SVM, $\vec{\theta} = \begin{pmatrix} \vec{w} \\ b \end{pmatrix}$, and we don't have $\vec{v}$ because we only have inequality constraints.

$$\mathcal{L}(\vec{w}, b, \vec{\lambda}) = \frac{1}{2}\|\vec{w}\|_2^2 + \sum_{i=1}^{N} \lambda_i \left(1 - y_i(\vec{w}^\top \vec{x}_i - b)\right)$$

The dual problem is

$$\max\limits_{\vec{\lambda}:\lambda_i \geq 0} \min\limits_{\vec{w},b} \mathcal{L}(\vec{w}, b, \vec{\lambda}) = \max\limits_{\vec{\lambda}:\lambda_i \geq 0} \min\limits_{\vec{w},b} \frac{1}{2}\|\vec{w}\|_2^2 + \sum_{i=1}^{N} \lambda_i \left(1 - y_i(\vec{w}^\top \vec{x}_i - b)\right)$$

Dual Form of the SVM

Let's solve the inner optimization problem of $\min_{\vec{w}, b} \mathcal{L}\left(\vec{w}, b, \vec{\lambda}\right)$:

$$\mathcal{L}\left(\vec{w}, b, \vec{\lambda}\right) = \frac{1}{2}\|\vec{w}\|_2^2 + \sum_{i=1}^{N} \lambda_i \left(1 - y_i(\vec{w}^\top \vec{x}_i - b)\right)$$

$$= \frac{1}{2}\vec{w}^\top\vec{w} + \sum_{i=1}^{N}\left(\lambda_i - \lambda_i y_i \vec{w}^\top \vec{x}_i + \lambda_i y_i b\right)$$

Recall: $\frac{\partial(\vec{x}^\top A \vec{x})}{\partial \vec{x}} = (A + A^\top)\vec{x}$ and $\frac{\partial(\vec{a}^\top \vec{x})}{\partial \vec{x}} = \vec{a}$

$$0 = \frac{\partial \mathcal{L}}{\partial \vec{w}} = \frac{1}{2}\left(I + I^\top\right)\vec{w} - \sum_{i=1}^{N}\lambda_i y_i \vec{x}_i = \frac{1}{2}(2I)\vec{w} - \sum_{i=1}^{N}\lambda_i y_i \vec{x}_i = \vec{w} - \sum_{i=1}^{N}\lambda_i y_i \vec{x}_i \Rightarrow \vec{w} = \sum_{i=1}^{N}\lambda_i y_i \vec{x}_i \quad (1)$$

$$0 = \frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{N}\lambda_i y_i \quad (2)$$

Both equations must hold at the critical point of $\mathcal{L}$.

Dual Form of the SVM

Let's check if the critical point is indeed a global minimum.

$$\frac{\partial \mathcal{L}}{\partial \vec{w}} = \vec{w} - \sum_{i=1}^{N} \lambda_i y_i \vec{x}_i \text{, and } \frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{N} \lambda_i y_i$$

Recall: $\frac{\partial (A\vec{x})}{\partial \vec{x}} = A^\top$

$$\frac{\partial^2 \mathcal{L}}{\partial \vec{w} \partial \vec{w}^\top} = I, \frac{\partial^2 \mathcal{L}}{\partial \vec{w} \partial b} = \vec{0}^\top, \frac{\partial^2 \mathcal{L}}{\partial b \partial \vec{w}^\top} = \vec{0}, \text{and } \frac{\partial^2 \mathcal{L}}{\partial b \partial b} = 0$$

Recall $\vec{\theta} = \begin{pmatrix} \vec{w} \\ b \end{pmatrix}$, so $\frac{\partial^2 \mathcal{L}}{\partial \vec{\theta} \partial \vec{\theta}^\top} = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial \vec{w} \partial \vec{w}^\top} & \frac{\partial^2 \mathcal{L}}{\partial b \partial \vec{w}^\top} \\ \frac{\partial^2 \mathcal{L}}{\partial \vec{w} \partial b} & \frac{\partial^2 \mathcal{L}}{\partial b \partial b} \end{pmatrix} = \begin{pmatrix} I & \vec{0} \\ \vec{0}^\top & 0 \end{pmatrix} \succeq 0$

So $\mathcal{L}$ is convex in $\vec{\theta}$, and so any critical point must be a global minimum.

Dual Form of the SVM

$$\mathcal{L}\left(\vec{w}, b, \vec{\lambda}\right) = \frac{1}{2} \vec{w}^\top \vec{w} + \sum_{i=1}^{N} \left(\lambda_i - \lambda_i y_i \vec{w}^\top \vec{x}_i + \lambda_i y_i b\right)$$

To find $\min_{\vec{w}, b} \mathcal{L}\left(\vec{w}, b, \vec{\lambda}\right)$, let's plug the equations that must hold at the critical point into the generalized Lagrangian.

Recall we have two equations that hold at the critical point: (1) $\vec{w} = \sum_{i=1}^{N} \lambda_i y_i \vec{x}_i$ and (2) $\sum_{i=1}^{N} \lambda_i y_i = 0$.

$$\min_{\vec{w}, b} \mathcal{L}\left(\vec{w}, b, \vec{\lambda}\right) = \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i \vec{x}_i\right)^\top \left(\sum_{i=1}^{N} \lambda_i y_i \vec{x}_i\right) + \sum_{i=1}^{N}\left(\lambda_i - \lambda_i y_i \left(\sum_{j=1}^{N} \lambda_j y_j \vec{x}_j\right)^\top \vec{x}_i + \lambda_i y_i b\right)$$

$$= \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i \vec{x}_i^\top\right)\left(\sum_{i=1}^{N} \lambda_i y_i \vec{x}_i\right) + \sum_{i=1}^{N}\left(\lambda_i - \lambda_i y_i \left(\sum_{j=1}^{N} \lambda_j y_j \vec{x}_j^\top\right) \vec{x}_i + \lambda_i y_i b\right)$$

Dual Form of the SVM

$$\min_{\vec{w},b} \mathcal{L}\left(\vec{w}, b, \vec{\lambda}\right) = \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i \vec{x}_i^\top\right)\left(\sum_{i=1}^{N} \lambda_i y_i \vec{x}_i\right) + \sum_{i=1}^{N}\left(\lambda_i - \lambda_i y_i \left(\sum_{j=1}^{N} \lambda_j y_j \vec{x}_j^\top\right)\vec{x}_i + \lambda_i y_i b\right)$$

$$= \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \vec{x}_i^\top \vec{x}_j\right) + \sum_{i=1}^{N}\left(\lambda_i - \lambda_i y_i \left(\sum_{j=1}^{N} \lambda_j y_j \vec{x}_j^\top\right)\vec{x}_i + \lambda_i y_i b\right)$$

$$= \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \vec{x}_i^\top \vec{x}_j\right) + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i y_i \left(\sum_{j=1}^{N} \lambda_j y_j \vec{x}_j^\top\right)\vec{x}_i + \sum_{i=1}^{N}\lambda_i y_i b$$

$$= \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \vec{x}_i^\top \vec{x}_j\right) + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i \lambda_j y_i y_j \vec{x}_j^\top \vec{x}_i + b\sum_{i=1}^{N}\lambda_i y_i$$

$$= \sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i \lambda_j y_i y_j \vec{x}_j^\top \vec{x}_i + b\sum_{i=1}^{N}\lambda_i y_i$$

Recall eq. (2) from a few slides ago: $0 = \frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{N} \lambda_i y_i$. Hence the last term is 0.

Dual Form of the SVM

Recall the dual problem is $\max\limits_{\vec{\lambda}:\lambda_i \geq 0 \; \forall i} \; \min\limits_{\vec{\theta}} \mathcal{L}\left(\vec{\theta}, \vec{\lambda}, \vec{v}\right)$.

From the previous slide,

$$\min_{\vec{w},b} \mathcal{L}\left(\vec{w}, b, \vec{\lambda}\right) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \vec{x}_j^{\top} \vec{x}_i \; \text{s.t.} \; \sum_{i=1}^{N} \lambda_i y_i = 0$$

So, the dual form of the SVM is:

$$\max_{\vec{\lambda}} \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \vec{x}_j^{\top} \vec{x}_i$$

subject to

$$\sum_{i=1}^{N} \lambda_i y_i = 0$$
$$\lambda_i \geq 0 \; \forall i$$

Primal vs. Dual of SVMs

Primal:

$$\underset{\vec{w},b}{\text{minimize}} \frac{1}{2}\|\vec{w}\|_2^2$$
$$\text{subject to } y_i\left(\vec{w}^\top \vec{x}_i - b\right) \geq 1 \; \forall i$$

Dual:

$$\underset{\vec{\lambda}}{\text{maximize}} \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \vec{x}_j^\top \vec{x}_i$$
$$\text{subject to } \sum_{i=1}^{N} \lambda_i y_i = 0$$
$$\lambda_i \geq 0 \; \forall i$$

Slater's Condition

Recall: Strong duality does not always hold.

One *sufficient* condition for strong duality to hold is *Slater's condition*.

Recall: We have a constrained optimization problem of the following form:

$$\min_{\vec{\theta}} f\left(\vec{\theta}\right) \text{ s.t. } g_i\left(\vec{\theta}\right) \leq 0 \;\forall i \in \{1, \dots, k\} \text{ and } h_i\left(\vec{\theta}\right) = 0 \;\forall i \in \{1, \dots, l\}$$

If $f\left(\vec{\theta}\right)$ and $g_i\left(\vec{\theta}\right)$ are convex in $\vec{\theta}$ $\forall i \in \{1, \dots, k\}$ and $h_i(\vec{\theta})$ is linear in $\vec{\theta}$ $\forall i$ $\in \{1, \dots, l\}$, and there exists a point $\vec{\theta}_0$ such that $g_i\left(\vec{\theta}\right) < 0 \;\forall i \in \{1, \dots, k\}$ and $h_i\left(\vec{\theta}\right)$ $= 0 \forall i \in \{1, \dots, l\}$, then strong duality holds.

Strict inequality!