Fundamentals of Linear Algebra and Optimization

Dual of the Hard Margin Support Vector Machine

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania
jean@cis.upenn.edu

May 7, 2020

Solving Hard Margin SVM Problem (SVM_{h2})

Recall the **Hard margin SVM** problem (SVM_{h2}):

$$\begin{aligned}
&\text{minimize} \quad \frac{1}{2}\left\|w\right\|^2, \qquad w \in \mathbb{R}^n \\
&\text{subject to} \\
&\qquad w^\top u_i - b \geq 1 \qquad i = 1, \ldots, p \\
&\qquad -w^\top v_j + b \geq 1 \qquad j = 1, \ldots, q.
\end{aligned}$$

Solving Hard Margin SVM Problem (SVM_{h2})

Recall the **Hard margin SVM** problem (SVM_{h2}):

$$
\begin{aligned}
&\text{minimize} \quad \frac{1}{2}\|w\|^2, \qquad w \in \mathbb{R}^n \\
&\text{subject to} \\
&\qquad\quad w^\top u_i - b \geq 1 \qquad i = 1, \ldots, p \\
&\qquad -w^\top v_j + b \geq 1 \qquad j = 1, \ldots, q.
\end{aligned}
$$

The main steps are the following.

Lagrangian of Hard Margin (SVM_{h2})

Step 1. Write the Lagrangian in matrix form.

Lagrangian of Hard Margin (SVM_{h2})

Step 1. Write the Lagrangian in matrix form.

Let X be the $n \times (p+q)$ matrix given by

$$X = \begin{pmatrix} -u_1 & \cdots & -u_p & v_1 & \cdots & v_q \end{pmatrix}.$$

Lagrangian of Hard Margin (SVM_{h2})

Step 1. Write the Lagrangian in matrix form.

Let X be the $n \times (p+q)$ matrix given by

$$X = \begin{pmatrix} -u_1 & \cdots & -u_p & v_1 & \cdots & v_q \end{pmatrix}.$$

We obtain the Lagrangian

$$L(w, b, \lambda, \mu) = \frac{1}{2} \begin{pmatrix} w^\top & b \end{pmatrix} \begin{pmatrix} I_n & 0_n \\ 0_n^\top & 0 \end{pmatrix} \begin{pmatrix} w \\ b \end{pmatrix} + \begin{pmatrix} w^\top & b \end{pmatrix} \begin{pmatrix} X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \\ \mathbf{1}_p^\top \lambda - \mathbf{1}_q^\top \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q}.$$

Dual Function of Hard Margin (SVM_{h2})

Step 2. Find the dual function $G(\lambda, \mu)$.

Dual Function of Hard Margin (SVM_{h2})

Step 2. Find the dual function $G(\lambda, \mu)$.

In order to find the dual function $G(\lambda, \mu)$, we need to minimize $L(w, b, \lambda, \mu)$ with respect to w and b and for this, since the objective function J is convex and since $\mathbb{R}^{n+1}$ is convex and open, a necessary and sufficient condition for a minimum is that $\nabla L_{w,b} = 0$, where $\nabla L_{w,b}$ is the gradient of $L(w, b, \lambda, \mu)$.

Dual Function of Hard Margin (SVM_{h2})

Step 2. Find the dual function $G(\lambda, \mu)$.

In order to find the dual function $G(\lambda, \mu)$, we need to minimize $L(w, b, \lambda, \mu)$ with respect to w and b and for this, since the objective function J is convex and since $\mathbb{R}^{n+1}$ is convex and open, a necessary and sufficient condition for a minimum is that $\nabla L_{w,b} = 0$, where $\nabla L_{w,b}$ is the gradient of $L(w, b, \lambda, \mu)$.

We have

$$\nabla L_{w,b} = \begin{pmatrix} w + X\begin{pmatrix}\lambda \\ \mu\end{pmatrix} \\ \mathbf{1}_p^\top \lambda - \mathbf{1}_q^\top \mu \end{pmatrix}.$$

Dual Function of Hard Margin (SVM_{h2})

The necessary and sufficient condition for a minimum is

$$\nabla L_{w,b} = 0,$$

Dual Function of Hard Margin (SVM_{h2})

The necessary and sufficient condition for a minimum is

$$\nabla L_{w,b} = 0,$$

which yields

$$w = -X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \qquad (*_1)$$

Dual Function of Hard Margin (SVM_{h2})

The necessary and sufficient condition for a minimum is

$$\nabla L_{w,b} = 0,$$

which yields

$$w = -X\begin{pmatrix}\lambda \\ \mu\end{pmatrix} \qquad (*_1)$$

and

$$\mathbf{1}_p^\top \lambda - \mathbf{1}_q^\top \mu = 0. \qquad (*_2)$$

Dual Function of Hard Margin (SVM_{h2})

The second equation can be written as

$$\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0. \qquad (*_3)$$

Dual Function of Hard Margin (SVM_{h2})

The second equation can be written as

$$\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0. \qquad (*_3)$$

Plugging back w from $(*_1)$ into the Lagrangian and using $(*_2)$ we get

$$G(\lambda, \mu) = -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q}, \qquad (*_4)$$

Dual Function of Hard Margin (SVM_{h2})

The second equation can be written as

$$\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0. \qquad (*_3)$$

Plugging back w from $(*_1)$ into the Lagrangian and using $(*_2)$ we get

$$G(\lambda, \mu) = -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q}, \qquad (*_4)$$

where $\begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q} = \sum_{i=1}^{p} \lambda_i + \sum_{j=1}^{q} \mu_j$.

Dual Function of Hard Margin (SVM_{h2})

Step 3. Write the dual as a minimization problem.

Dual Function of Hard Margin (SVM_{h2})

Step 3. Write the dual as a minimization problem.

Maximizing the dual function $G(\lambda, \mu)$ over its domain of definition is equivalent to maximizing

$$\widehat{G}(\lambda, \mu) = -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q}$$

Dual Function of Hard Margin (SVM_{h2})

Step 3. Write the dual as a minimization problem.

Maximizing the dual function $G(\lambda, \mu)$ over its domain of definition is equivalent to maximizing

$$\widehat{G}(\lambda, \mu) = -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q}$$

subject to the constraint

$$\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0,$$

Convert Dual to a Minimization Problem

so we formulate the dual program as,

$$\text{maximize} \quad -\frac{1}{2}\begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix} X^\top X \begin{pmatrix}\lambda \\ \mu\end{pmatrix} + \begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix} \mathbf{1}_{p+q}$$

$$\text{subject to}$$

$$\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0$$

$$\lambda \geq 0, \; \mu \geq 0,$$

Dual Function of Hard Margin (SVM_{h2})

or equivalently, **Dual of the Hard margin SVM** (SVM_{h2}):

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} - \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q} \\ &\text{subject to} \\ &\qquad \sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0 \\ &\qquad \lambda \geq 0, \ \mu \geq 0. \end{aligned}$$

Solving the Dual Program of (SVM_{h2})

Step 4. Solve the dual program.

Solving the Dual Program of (SVM_{h2})

Step 4. Solve the dual program.

This step involves using numerical procedures typically based on gradient descent to find λ and μ, for example, ADMM.

Solving the Dual Program of (SVM_{h2})

Step 4. Solve the dual program.

This step involves using numerical procedures typically based on gradient descent to find λ and μ, for example, ADMM.

Once λ and μ are determined, w is determined by $(*_1)$, namely

$$w = -X \begin{pmatrix} \lambda \\ \mu \end{pmatrix}.$$

Solving the Dual Program of (SVM_{h2})

Step 4. Solve the dual program.

This step involves using numerical procedures typically based on gradient descent to find λ and μ, for example, ADMM.

Once λ and μ are determined, w is determined by $(*_1)$, namely

$$w = -X \begin{pmatrix} \lambda \\ \mu \end{pmatrix}.$$

To determine b we use the KKT conditions.

Using the KKT Conditions of (SVM_{h2})

Because the primal always has a solution, so does the dual, which implies that there is at least some i_0 such that $\lambda_{i_0} > 0$. But then the constraint $\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0$ implies that there is also some j_0 such that $\mu_{j_0} > 0$.

Using the KKT Conditions of (SVM_{h2})

Because the primal always has a solution, so does the dual, which implies that there is at least some i_0 such that $\lambda_{i_0} > 0$. But then the constraint $\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0$ implies that there is also some j_0 such that $\mu_{j_0} > 0$.

By the KKT conditions, since the corresponding constraints are active, we have

$$w^\top u_{i_0} - b = 1, \quad -w^\top v_{j_0} + b = 1,$$

Using the KKT Conditions of (SVM_{h2})

Because the primal always has a solution, so does the dual, which implies that there is at least some i_0 such that $\lambda_{i_0} > 0$. But then the constraint $\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0$ implies that there is also some j_0 such that $\mu_{j_0} > 0$.

By the KKT conditions, since the corresponding constraints are active, we have

$$w^\top u_{i_0} - b = 1, \quad -w^\top v_{j_0} + b = 1,$$

so we obtain

$$b = w^\top (u_{i_0} + v_{j_0})/2.$$

Using the KKT Conditions of (SVM_{h2})

Because the primal always has a solution, so does the dual, which implies that there is at least some i_0 such that $\lambda_{i_0} > 0$. But then the constraint $\sum_{i=1}^{p} \lambda_i - \sum_{j=1}^{q} \mu_j = 0$ implies that there is also some j_0 such that $\mu_{j_0} > 0$.

By the KKT conditions, since the corresponding constraints are active, we have

$$w^\top u_{i_0} - b = 1, \quad -w^\top v_{j_0} + b = 1,$$

so we obtain

$$b = w^\top (u_{i_0} + v_{j_0})/2.$$

The *support vectors* are those for which the constraints are active.

Averaging Over Indices

For improved numerical stability, we can average over the sets of indices defined as $I_{\lambda>0} = \{i \in \{1, \ldots, p\} \mid \lambda_i > 0\}$ and $I_{\mu>0} = \{j \in \{1, \ldots, q\} \mid \mu_j > 0\}$.

Averaging Over Indices

For improved numerical stability, we can average over the sets of indices defined as $I_{\lambda>0} = \{i \in \{1, \ldots, p\} \mid \lambda_i > 0\}$ and $I_{\mu>0} = \{j \in \{1, \ldots, q\} \mid \mu_j > 0\}$.

We obtain

$$b = w^\top \left(\left(\sum_{i \in I_{\lambda>0}} u_i \right) / |I_{\lambda>0}| + \left(\sum_{j \in I_{\mu>0}} v_j \right) / |I_{\mu>0}| \right) / 2.$$