Module 12

Nonlinear Optimization and Hard Margin SVM Solutions

Problem 1: 50 points total

Linear programming with box constraints is the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax = b \\ & l \leq x \leq u, \end{aligned}$$

where A is an $m \times n$ matrix, $c, u, l, x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, with $l \leq u$ (which means that $l_i \leq u_i$, for $i = 1, \ldots, n$).

(1) (20 points) Prove that the dual of the above program is the following program:

$$\begin{aligned} \text{maximize} \quad & -\nu^\top b - \lambda_1^\top u + \lambda_2^\top l \\ \text{subject to} \quad & A^\top \nu + \lambda_1 - \lambda_2 + c = 0 \\ & \lambda_1 \geq 0, \quad \lambda_2 \geq 0. \end{aligned}$$

(2) (10 points) The primal problem in (1) can be reformulated by incorporating the constraints $l \leq x \leq u$ into the objective function by defining

$$f_0(x) = \begin{cases} c^\top x & \text{if } l \leq x \leq u \\ +\infty & \text{otherwise.} \end{cases}$$

The primal is reformulated as

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax = b. \end{aligned}$$

Prove that the new dual function is given by

$$G(\nu) = \inf_{l \leq x \leq u} (c^\top x + \nu^\top (Ax - b)).$$

(3) (20 points) Given any real number $s \in \mathbb{R}$, let

$$s^+ = \max\{s, 0\}, \quad s^- = \max\{-s, 0\}.$$

Prove that for any reals $s, \lambda, \mu \in \mathbb{R}$ with $\lambda \leq \mu$,

$$\inf_{\lambda \leq y \leq \mu} sy = \lambda s^+ - \mu s^-.$$

Hint. Consider the cases $s \geq 0$ and $s \leq 0$.

We extend the above operators to vectors $z \in \mathbb{R}^n$ componentwise by

$$z^+ = (z_1^+, \ldots . z_n^+), \quad z^- = (z_1^-, \ldots . z_n^-).$$

For any $w \in \mathbb{R}^n$, prove that

$$\inf_{l \leq x \leq u} x^\top w = l^\top w^+ - u^\top w^-.$$

Use the above to prove that

$$G(\nu) = -\nu^\top b + l^\top (A^\top \nu + c)^+ - u^\top (A^\top \nu + c)^-$$

and deduce that the dual program is the unconstained problem

$$\text{maximize} \quad -\nu^\top b + l^\top (A^\top \nu + c)^+ - u^\top (A^\top \nu + c)^-$$

with respect to ν.

Solution. (1) First we rewrite the inequalities $x \leq u$ and $l \leq x$ as $x - u \leq 0$ and $-x + l \leq 0$. To form the Lagrangian we assign the Lagrange multipliers $\nu \in \mathbb{R}^m$ to the system of equations $Ax - b = 0$, $\lambda_1 \in \mathbb{R}^n_+$ ($\lambda_1 \geq 0$) to the inequalities $x - u \leq 0$, and $\lambda_2 \in \mathbb{R}^n_+$ ($\lambda_2 \geq 0$) to the inequalities $-x + l \leq 0$. The Lagrangian $L(x, \nu, \lambda_1, \lambda_2)$ is given by

$$L(x, \nu, \lambda_1, \lambda_2) = c^\top x + \nu^\top (Ax - b) + \lambda_1^\top (x - u) + \lambda_2^\top (-x + l).$$

The Lagrangian can be rewritten as

$$L(x, \nu, \lambda_1, \lambda_2) = (c^\top + \nu^\top A + \lambda_1^\top - \lambda_2^\top) x - \nu^\top b - \lambda_1^\top u + \lambda_2^\top l.$$

Next we want to contruct the dual function $G(\nu, \lambda_1, \lambda_2)$ obtained by minimizing $L(x, \nu, \lambda_1, \lambda_2)$ over $x \in \mathbb{R}^n$,

$$G(\nu, \lambda_1, \lambda_2) = \inf_{x \in \mathbb{R}^n} L(x, \nu, \lambda_1, \lambda_2).$$

The linear form $L(x, \nu, \lambda_1, \lambda_2)$ (in x, with $\nu, \lambda_1, \lambda_2$ fixed) is unbounded below unless

$$c^\top + \nu^\top A + \lambda_1^\top - \lambda_2^\top = 0,$$

which (by transposition) is equivalent to

$$A^\top \nu + \lambda_1 - \lambda_2 + c = 0.$$

Therefore the dual function $G(\nu, \lambda_1, \lambda_2)$ is given by

$$G(\nu, \lambda_1, \lambda_2) = \begin{cases} -\nu^\top b - \lambda_1^\top u + \lambda_2^\top l & \text{if } A^\top \nu + \lambda_1 - \lambda_2 + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The above defintion yields the dual program

$$\begin{aligned} &\text{maximize} && -\nu^\top b - \lambda_1^\top u + \lambda_2^\top l \\ &\text{subject to} && A^\top \nu + \lambda_1 - \lambda_2 + c = 0 \\ &&& \lambda_1 \geq 0, \quad \lambda_2 \geq 0. \end{aligned}$$

(2) This time we only need to assign the Lagrange multipliers $\nu \in \mathbb{R}^m$ to the system of equations $Ax - b = 0$ and we obtain the Lagrangian

$$L(x, \nu) = f_0(x) + \nu^\top (Ax - b).$$

The dual function $G(\nu)$ is obtained by minimizing $L(x, \nu)$ over $x \in \mathbb{R}^n$ (with ν fixed). By definition of

$$f_0(x) = \begin{cases} c^\top x & \text{if } l \leq x \leq u \\ +\infty & \text{otherwise,} \end{cases}$$

this minimization can be restricted to those $x \in \mathbb{R}^n$ such that $l \leq x \leq u$, in which case $f_0(x) = c^\top x$, so

$$G(\nu) = \inf_{l \leq x \leq u} (c^\top x + \nu^\top (Ax - b)).$$

(3) We have

$$c^\top x + \nu^\top (Ax - b) = -\nu^\top b + (\nu^\top A + c^\top) x = -\nu^\top b + x^\top (A^\top \nu + c),$$

and so

$$G(\nu) = \inf_{l \leq x \leq u} (-\nu^\top b + x^\top (A^\top \nu + c)).$$

Given any real number $s \in \mathbb{R}$, let

$$s^+ = \max\{s, 0\}, \quad s^- = \max\{-s, 0\}.$$

Then for any fixed reals $s, \lambda, \mu \in \mathbb{R}$ with $\lambda \leq \mu$, we claim that

$$\inf_{\lambda \leq y \leq \mu} sy = \lambda s^+ - \mu s^-.$$

Case 1. $s \geq 0$. In this case $s^+ = s$, and since $-s \leq 0$, we have $s^- = 0$. The minimum of $y \mapsto sy$ for $\lambda \leq y \leq \mu$ is $\lambda s = \lambda s^+$.

Case 2. $s \leq 0$. In this case $s^+ = 0$ and $s^- = -s \geq 0$. The minimum of $y \mapsto sy$ for $\lambda \leq y \leq \mu$ is $\mu s = -\mu s^-$.

In summary, we have

$$\inf_{\lambda \leq y \leq \mu} sy = \lambda s^+ - \mu s^-,$$

as claimed.

For any $w \in \mathbb{R}^n$, the minimum of a linear form $x \mapsto x^\top w = \sum_{i=1}^n x_i w_i$ for $l \leq x \leq u$ is obtained by computing the minima componentwise,

$$\begin{aligned}\inf_{l \leq x \leq u} \left(\sum_{i=1}^n w_i x_i \right) &= \sum_{i=1}^n \inf_{l_i \leq x_i \leq u_i} w_i x_i \\ &= \sum_{i=1}^n (l_i w_i^+ - u_i w_i^-) \\ &= \sum_{i=1}^n l_i w_i^+ - \sum_{i=1}^n u_i w_i^- \\ &= l^\top w^+ - u^\top w^-,\end{aligned}$$

so we get

$$\inf_{l \leq x \leq u} x^\top w = l^\top w^+ - u^\top w^-.$$

As a consequence, since

$$G(\nu) = \inf_{l \leq x \leq u} (-\nu^\top b + x^\top (A^\top \nu + c)),$$

with $w = A^\top \nu + c$ we obtain

$$G(\nu) = -\nu^\top b + l^\top (A^\top \nu + c)^+ - u^\top (A^\top \nu + c)^-.$$

If follows immediately that the dual program is the unconstrained problem

$$\text{maximize} \quad -\nu^\top b + l^\top (A^\top \nu + c)^+ - u^\top (A^\top \nu + c)^-$$

with respect to ν.

Total: 50 points