Ass4 Solutions

Chang Li

September 18, 2016

1 Solutions

1.1 Prob 6.2

1.1.1 l2 norm

We want the closest point $x1$ in the subspace to b, namely we want the error $e = x1 - b$ to be orthogonal to $C(1)$, which equivalent to being in $N(1^T)$. The problem reaches it's minimum when solution satisfies:

$$1^T 1x = 1^T b$$

Suppose $b \in R^n$ then $x = \frac{1^T b}{n}$.

1.1.2 l1 norm

The equivalent LP problem is:

$$\text{minimize} \sum_{i=1}^{n} 1^T y$$

$$\text{s.t.} \ -y \preceq x1 - b \preceq y$$

This is equivalent to let $y_i = |x - b_i|$. Therefore the problem reaches minimum when x equals the median of the vector b.

1.1.3 l∞ norm

The equivalent LP problem is:

$$\text{minimize} \quad \sum_{i=1}^{n} 1^T y$$

$$\text{s.t.} \ -y1 \preceq x1 - b \preceq y1$$

This is equivalent to let $y \geq \max |x - b_i|$. Therefore the problem reaches its minimum when y equals the midpoint of vector b. Suppose $b_1 \ldots b_n$ is written in nondecreasing order, then $y = \frac{b_n - b_1}{2}$.

1.2 Prob 6.6

Because the variables are x and r, the Lagrangian of the original problem is:

$$L(x, r, v) = \sum_{i=1}^{m} \phi(r_i) + v^T(Ax - b - r)$$

Therefore, L reaches minimum when $v^T A = 0$. Plug L into $g(v)$ we have:

$$g(v) = \begin{cases} -b^t v + \sum_{i=1}^m \inf_{r_i}(\phi(r_i) - v_i r_i) & A^T v = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Because

$$\inf(\phi(r_i) - v_i r_i) = -\sup(v_i r_i - \phi(r_i)) = -\phi^*(v_i)$$

Where $\phi^*(v_i)$ is the conjugate function of $\phi(r_i)$. Therefore the dual problem can be written as:

$$\text{maximize} \quad -b^T v - \sum_{i=1}^{m} \phi^*(v_i)$$
$$\text{s.t.} \;\; A^T v = 0$$

To find the dual problem of penalty functions we only need to find their conjugate functions.

1.2.1 a

We first verify the domain of the conjugate function ϕ^*. We can write $yx - \phi(x)$ as:

$$\phi^*(y) = \begin{cases} (y+1)x + 1 & x < -1 \\ yx & -1 \leq x \leq 1 \\ (y-1)x + 1 & x > 1 \end{cases}$$

Then it's clear that when $|y| > 1$ it is not in the domain of ϕ^*. This is because if $y < -1$, let $x \to -\infty$ and if $y > 1$ then let $x \to \infty$. The value is unbounded above.

When $|y| \leq 1$, the value is always less than 0 if $|x| > 1$. When $|x| \leq 1$, the value reaches its maximum $|y|$ by setting $x = -1$ if $y \leq 0$ and $x = 1$ if $y > 0$. Therefore, The conjugate function of deadzone-linear penalty is:

$$\phi^*(z) = \begin{cases} |z| & |z| \leq 1 \\ \infty & \text{otherwise} \end{cases}$$

To plug $\phi^*(z)$ into the object function, we need to ensure the function is bounded. Therefore the vector v has to satisfy the constraint $||v||_\infty \leq 1$. Then we have:

$$\text{maximize} \quad -b^T v - \sum_{i=1}^{m} |v_i|$$
$$\text{s.t.} \;\; A^T v = 0, \;\; ||v||_\infty \leq 1$$

1.2.2 b

We first verify the domain of the conjugate function ϕ^*. We can write $yx - \phi(x)$ as:

$$\phi^*(y) = \begin{cases} (y+2)x + 1 & x < -1 \\ yx - x^2 & -1 \leq x \leq 1 \\ (y-2)x + 1 & x > 1 \end{cases}$$

Then it's clear that when $|y| > 2$ it is not in the domain of ϕ^*. This is because if $y < -2$, let $x \to -\infty$ and if $y > 2$ then let $x \to \infty$. The value is unbounded above.

When $|y| \leq 2$, the value is always less than 0 if $|x| > 1$. When $|x| \leq 1$, the value reaches its maximum $\frac{y^2}{4}$ by setting $x = \frac{y}{2}$. Therefore, The conjugate function of huber penalty is:

$$\phi^*(z) = \begin{cases} \frac{z^2}{4} & |z| \leq 2 \\ \infty & \text{otherwise} \end{cases}$$

To plug $\phi^*(z)$ into the object function, we need to ensure the function is bounded. Therefore the vector v has to satisfy the constraint $||v||_\infty \leq 2$. Then we have:

$$\text{maximize} \quad -b^T v - \frac{1}{4}\sum_{i=1}^{m} z_i^2$$

$$\text{s.t.} \;\; A^T v = 0, \;\; ||v||_\infty \leq 2$$

1.3 Prob 7.6

Because $h(y) = x = ay - b$, so the density of y is $p(y) = p(h(y))h'(y)$:

$$p(y) = ap(ay - b)$$

Then the log-likelihood becomes:

$$\log p(y) = \log a + \log p(ay - b)$$

Given samples $y_1 \ldots y_n$ of y, the ML estimate of a and b is equivalent to maximize the sum of log-likelihood:

$$\sum_{i=1}^{n} \log p(y_i) = n\log a + \sum_{i=1}^{n} \log p(ay_i - b)$$

Since p is log concave, then the function is also a concave function. Therefore, the ML estimates of a and b is a concave(its negative is convex) problem.

When $p(x) = e^{-2|x|}$, the ML estimation becomes solving

$$\text{minimize} \quad -n\log a + 2\sum_{i=1}^{n} |ay_i - b|$$

Let $c = b/a$, the problem reaches minimum when c equals the median of y (proved by section 1.1.1).

$$\text{minimize} \quad -n\log a + 2a\sum_{i=1}^{n}|y_i - c|$$

Suppose the median of y is y_k. Then a^* can be found by setting the partial derivative of function $-n\log a + 2\sum_{i=1}^{n}|ay_i - b|$ $\frac{\partial f}{\partial a} = 0$. Namely:

$$a^* = \frac{n}{2\sum_{i=1}^{n}|y_i - c|}$$

Therefore, $b^* = y_k a^*$.

1.4 Prob 8.24

Because ρ is the upper bound of $||u||_2$ and any vector has a greater norm will leads to failing of separating. So that $u^T x_i$ gets its minimum value when u is on the same line as x_i but with the opposite direction.

Therefore, for all $||u||_2 \leq \rho$, we have $-\rho||x_i||_2 \leq u^T x_i \leq \rho||x_i||_2$. Therefore, we can rewrite the conditions for weight error margin as for $i = 1 \ldots N$ and $j = 1 \ldots M$:

$$a^T x_i - \rho||x_i||_2 \geq b_i, \quad a^T y_j + \rho||y_j||_2 \leq b_j$$

Therefore we have:

$$\rho = \min\left(\min_{i=1\ldots N} \frac{a^T x_i - b_i}{||x_i||_2}, \min_{j=1\ldots M} \frac{b_j - a^T y_j}{||y_j||_2}\right)$$

To maximize ρ, we can rewrite the optimizing problem of a, b and t into the maximizing problem:

$$\text{maximize} \;\; t$$

$$s.t. \begin{cases} a^T x_i - b_i \geq t||x_i||_2 & i = 1\ldots N \\ b_j - a^T y_j \geq t||y_j||_2 & j = 1\ldots M \\ ||a||_2 \leq 1 \end{cases}$$

Bibliography