Convex Optimization
HW3

Q1: the Lasso problem is defined as follows

$$\min_{w} \frac{1}{2}\|Xw-y\|_2^2+\lambda\|w\|_1$$

where $X=(x_1^T, \dots x_n^T)\in\mathbb{R}^{n\times d}$, $y=(y_1, \dots y_n)\in\mathbb{R}^n$, $w\in\mathbb{R}^d$ and $\lambda>0$

• we perform change of variable so the problem is equivalent to

$$\min_{z,w} \frac{1}{2}\|z\|_2^2+\lambda\|w\|_1$$
$$Xw-y-z=0$$

the Lagrangian is:

$$L(z,w,v)=\frac{1}{2}\|z\|_2^2+\lambda\|w\|_1+v^T(Xw-y-z)$$
$$=\left(\frac{1}{2}\|z\|_2^2-v^Tz\right)+\left(\lambda\|w\|_1+v^TXw\right)+v^Ty$$

so: $g(v)=\inf_{w,z} L(z,w,v)=-\sup_{w,z}-L(z,w,v)$

$$=\underbrace{-\sup_{z}\left(v^Tz-\frac{1}{2}\|z\|_2^2\right)}_{(1)}\underbrace{-\sup_{w}\left(-\lambda\|w\|_1+v^TXw\right)}_{(2)}+v^Ty$$

(1) let $h(z)=v^Tz-\frac{1}{2}\|z\|_2^2$ we see that this function is concave so can be maximized by setting $\nabla h(z)=0$

$$\nabla h(z)=v-z=0 \Rightarrow z^*=v$$

$$h^*(z)=\frac{1}{2}\|v\|_2^2$$

(2) $\sup_w \left(-\lambda \|w_1\| - v^T X w\right) = \lambda \sup_w \left(-\frac{v^T X}{\lambda} w - \|w\|_1\right)$

$= \lambda \sup_w \left(-\left(\frac{X^T v}{\lambda}\right)^T w - \|w\|_1\right)$

and we have seen that

$\|\cdot\|_1^*(y) = \begin{cases} 0 & \text{if } \|y\|_\infty < 1 \\ \infty & \text{elsewhere} \end{cases}$

so $\sup \left(-\left(\frac{X^T v}{\lambda}\right)^T w - \|w\|_1\right) = \begin{cases} 0 & \text{if } \|X^T v\|_\infty < \lambda \\ \infty & \text{elsewhere} \end{cases}$

then

$g(v) = -\left(\frac{1}{2}\|v\|_2^2 + y^T v\right) + \|\cdot\|_1^*\left(\frac{X^T v}{\lambda}\right)$

so the dual problem is

$$\max \; -\frac{1}{2}\|v\|_2^2 - y^T v$$
$$\text{st} \quad \|X^T v\|_\infty \leq \lambda$$

so to get the formulation :

$$\min \; v^T Q v + p^T v$$
$$\text{st} \quad Av \leq b$$

we set : $Q = \frac{1}{2} I_n$

$p = y$

$A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}$

& $b = \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}$

Scanné avec CamScanner

2) • Implement the Newton method to solve the centering step

$$\min f(v) = t(v^T Q v + p^T v) + \sum_{i=1}^{2d} \log(b_i - A_i v)$$

where $A = \begin{bmatrix} -A_1- \\ \vdots \\ -A_n- \end{bmatrix}$

we compute the gradient and the Hessian:

$$\nabla f(v) = t(2Qv + p) + \sum_{i=1}^{2d} \frac{A_i^T}{b_i - A_i v}$$

$$\nabla^2 f(v) = t(2Q) + \sum_{i=1}^{2d} \frac{A_i^T A_i}{(b_i - A_i v)^2}$$