Convex Optimization - Homework 3

Petrovich Mathis

Question 1

$$\begin{aligned}
&\underset{w}{\text{minimize}}\ \frac{1}{2}\left\|Xw-y\right\|_2^2+\lambda\left\|w\right\|_1 \quad \text{(LASSO)}\\
\Leftrightarrow\ &\underset{w,z}{\text{minimize}}\ \frac{1}{2}\left\|z-y\right\|_2^2+\lambda\left\|w\right\|_1\\
&\text{subject to } z=Xw
\end{aligned}$$

- Lagragian:

$$\begin{aligned}
\mathcal{L}(w,z,u) &= \frac{1}{2}\left\|z-y\right\|_2^2+\lambda\left\|w\right\|_1+u^T(z-Xw)\\
&= \frac{1}{2}\left\|z-y\right\|_2^2+u^Tz+\lambda\left\|w\right\|_1-u^TXw
\end{aligned}$$

- Dual problem: $g(v)=\min_{w,z}\mathcal{L}(w,z,v)$

$\mathcal{L}$ is convex in z, $\nabla_z\mathcal{L}(w,z^*,v)=0\Rightarrow z^*-y+v=0\Rightarrow z^*=y-v$

$$\begin{aligned}
g(v) &= \min_w\left[\frac{1}{2}v^Tv+v^Ty-v^Tv+\lambda\left\|w\right\|_1-v^TXw\right]\\
g(v) &= -\frac{1}{2}v^Tv+v^Ty+\min_w\left[\lambda\left\|w\right\|_1-v^TXw\right]
\end{aligned}$$

- Let's focus on this term: $T=\min_w\left[\lambda\left\|w\right\|_1-v^TXw\right]=-\lambda\left\|.\right\|_1^*(\frac{X^Tv}{\lambda})$

with $\left\|.\right\|_1^*(x)=\sup_w(x^Tw-\left\|w\right\|_1)$

Let $x\in\mathbb{R}^d$

- If $\exists i;|x_i|>1$ then let $w=t*\delta_i$ with δ_i the zero vector but with a 1 in index i and $t\in\mathbb{R}$. Then $x^Tw-\left\|w\right\|_1=t*x_i-|t|=(|x_i|-1)*|t|$ (if we take t with the sign of x_i). With $t\to\pm\infty$ we get $\left\|.\right\|_1^*(x)=+\infty$.
- Else $\forall i;|x_i|<1$ so $|x_iw_i|<|w_i|\ \forall i$ so $x^Tw-\left\|w\right\|_1\leq 0$ (equality when $x=0$)

So $\left\|.\right\|_1^*(v)=\sup_w(v^Tw-\left\|w\right\|_1)$

$$\left\|.\right\|_1^*(x)=\begin{cases}0 & \text{if} \quad \forall i\ |x_i|<1\\ +\infty & \text{otherwise}\end{cases}$$

So we have

$$T = \begin{cases} 0 & \text{if} \quad X^T v \preceq \lambda \mathbb{1} \text{ and } -X^T v \preceq \lambda \mathbb{1} \\ -\infty & \text{otherwise} \end{cases}$$

Thus, because we want to maximize g for the dual problem, we need the constraint to put T to 0:

$$\begin{aligned} &\underset{v}{\text{maximize}} \; -\frac{1}{2}v^T v + v^T y \\ &\quad \text{subject to } X^T v \preceq \lambda \mathbb{1} \text{ and } -X^T v \preceq \lambda \mathbb{1} \\ \Leftrightarrow &\underset{v}{\text{minimize}} \; \frac{1}{2}v^T v - v^T y \\ &\quad \text{subject to } X^T v \preceq \lambda \mathbb{1} \text{ and } -X^T v \preceq \lambda \mathbb{1} \end{aligned}$$

Let $A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}$, $p = -y$, $b = \lambda \mathbb{1}$ and $Q = \frac{1}{2} Id$ ($Q \succeq 0$)

This problem is equivalent to:

$$\begin{aligned} &\underset{v}{\text{minimize}} \; v^T Q v + p^T v \quad \text{(QP)} \\ &\text{subject to } A^T v \preceq b \end{aligned}$$

Question 2

Let compute the dual of our QP problem.

$$\mathcal{L}(v, \alpha_1, \alpha_2) = \frac{1}{2}v^T v - v^T y + \alpha_1^T (X^T v - \lambda) + \alpha_2^T (-X^T v - \lambda)$$

$\mathcal{L}$ is convex in v, $\nabla_v \mathcal{L}(v^*, \alpha_1, \alpha_2) = 0 = v^* - y + X\alpha_1 - X\alpha_2 \Rightarrow v^* = y - X(\alpha_2 - \alpha_1)$
Let $w = (\alpha_1 - \alpha_2)$

$$\begin{aligned} g(\alpha_1, \alpha_2) &= \frac{1}{2} \|Xw - y\|_2^2 - (y - Xw)^T y + w^T (X^T (y - Xw)) - \lambda(\alpha_1 + \alpha_2)^T \mathbb{1} \\ &= -\frac{1}{2} \|Xw - y\|_2^2 - \lambda(\alpha_1 + \alpha_2)^T \mathbb{1} \end{aligned}$$

The dual is :

$$\begin{aligned} &\underset{w, \alpha_1, \alpha_2}{\text{minimize}} \; \frac{1}{2} \|Xw - y\|_2^2 + \lambda(\alpha_1 + \alpha_2)^T \mathbb{1} \\ &\text{subject to } \alpha_1 \succeq 0, \alpha_2 \succeq 0 \end{aligned}$$

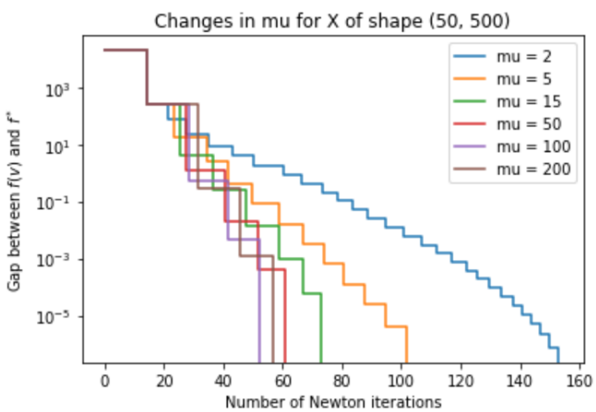
We can notice that for an optimal solution $w^* = \alpha_1^* - \alpha_2^*$ of this problem, $\forall i$, $\alpha_{1,i}^*$ or $\alpha_{2,i}^*$ should be zero. Indeed, if $\alpha_{1,i}^* - \alpha_{2,i}^* \geq 0$ then we can take $new\alpha_{1,i}^* = \alpha_{1,i}^* - \alpha_{2,i}^*$ and $new\alpha_{2,i}^* = 0$. This doesn't change w but reduce $\lambda(\alpha_1 + \alpha_2)^T \mathbb{1}$. (The case $\alpha_{1,i}^* - \alpha_{2,i}^* \leq 0$ is similar).

With this result, we can deduce than $\alpha_1 + \alpha_2 = |w|$. So the dual of our (QP) is equivalent to:

$$\underset{w}{\text{minimize}} \ \frac{1}{2} \|Xw - y\|_2^2 + \lambda \|w\|_1 \quad \text{(LASSO)}$$

This result is usefull to recover w from the resolution of (QP). Indeed, with the barrier method, we have acces to the dual variable (α_1 and α_2), so we have acces to w.
With approximate complementary slackness, $\alpha_1 = \frac{1}{-t(X^T v - \lambda)}$ and $\alpha_2 = \frac{1}{-t(-X^T v - \lambda)}$

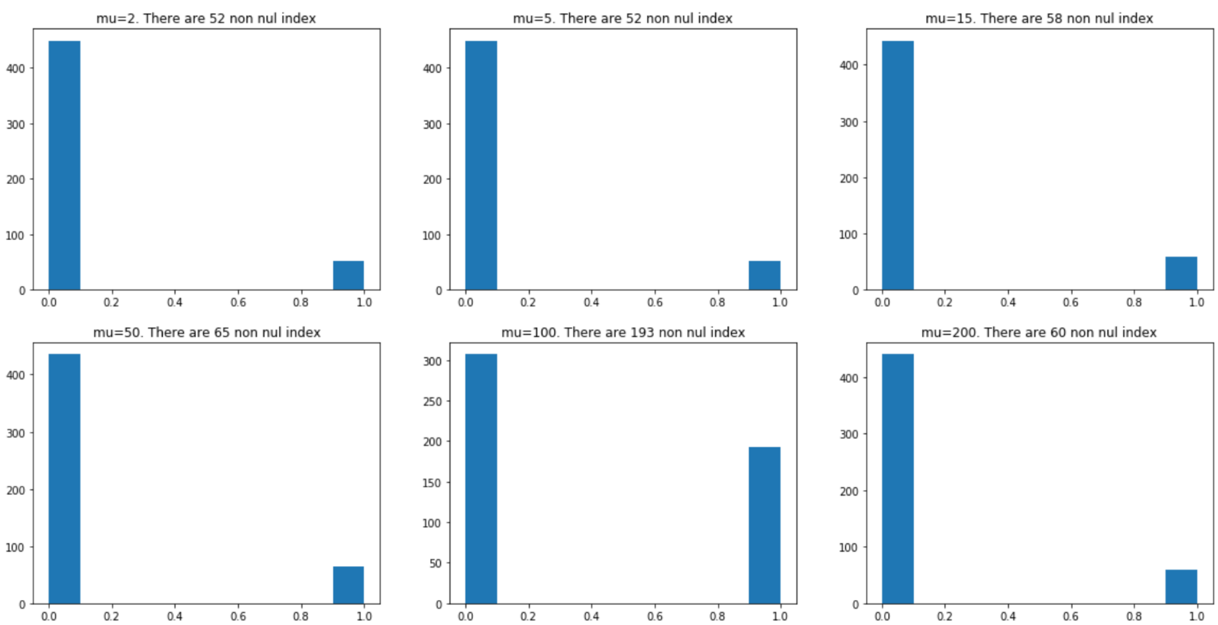
Results



We can notice that when mu become larger, there are less outer iterations but there are more inner iterations (centering step with newton).

With this plot, we can see that $mu = 100$ is the mu when there are the minimim of newton iterations.

To see the inpact of w, I made histograms of w and I count how many index are very close to 0.



We can notice that in general when mu become larger, there are less zero in w (less sparse).
$mu = 50$ seems to be an acceptable choice for mu.