MVA / Convex Optimization 2022 / HW3

Imane Elbacha

28 November, 2022

Given $x_1, \ldots, x_n \in \mathbb{R}^d$ data vectors and $y_1, \ldots, y_n \in \mathbb{R}$ observations, we are searching for regression parameters $w \in \mathbb{R}^d$ which fit data inputs to observations y by minimizing their squared difference. In a high dimensional setting (when $n \ll d$) a ℓ_1 norm penalty is often used on the regression coefficients w in order to enforce sparsity of the solution (so that w will only have a few non-zeros entries). Such penalization has well known statistical properties, and makes the model both more interpretable, and faster at test time. From an optimization point of view we want to solve the following problem called LASSO (which stands for Least Absolute Shrinkage Operator and Selection Operator)

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

in the variable $w \in \mathbb{R}^d$, where $X = \left(x_1^T, \ldots, x_n^T\right) \in \mathbb{R}^{n\times d}, y = (y_1, \ldots y_n) \in \mathbb{R}^n$ and $\lambda > 0$ is a regularization parameter.

1.Derive the dual problem of LASSO and format it as a general Quadratic Problem as follows:

$$\begin{aligned} &\min v^T Q v + p^T v \\ \text{s.t. } &Av \preceq b \end{aligned}$$

in variable $v \in \mathbb{R}^n$, where $Q \succeq 0$.

Solution: We can rewrite the problem as follows:

$$\min_w \frac{1}{2}\|z\|_2^2 + \lambda\|w\|_1 \text{ with } \quad z = Xw - y$$

The lagrangian of the new problem is: Let $w \in \mathbb{R}^d, z \in \mathbb{R}^n, \nu \in \mathbb{R}^n$

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

and the dual function is

$$\begin{aligned} g(\nu) &= \inf_{w,z} \mathcal{L}(w, z, \nu) \\ &= y^T\nu + \underbrace{\inf_z \left(\frac{1}{2}\|z\|_2^2 + \nu^T z\right)}_{1} + \underbrace{\inf_w \left(\lambda\|w\|_1 - \left(X^T\nu\right)^T w\right)}_{2} \end{aligned}$$

For part 1 of the equation : The function $h : z \mapsto \frac{1}{2}\|z\|_2^2 + \nu^T z$ is convex and differentiable.
We compute the gradient: $\nabla h(z) = z + \nu$ and we have $\nabla h(z) = 0$ iff $z = -\nu$.
Hence, the minimum of h is $-\frac{1}{2}\|\nu\|_2^2$.
For part 2 of the equation :
We already proved in HW2 Ex 2.1: $u \in \mathbb{R}^n$,

$$\|\cdot\|_1^*(u) = \begin{cases} 0 & \text{if } \|u\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

With this result we can compute the conjugate of the second part :

We have:

$$\inf_w \lambda\|w\|_1 - \left(X^T\nu\right)^T w = \sup_w \left(\frac{1}{\lambda}X^T\nu\right)^T w - \|w\|_1 = \|\cdot\|_1^*\left(\frac{1}{\lambda}X^T\nu\right)$$

Hence we can write the dual function:

$$g(\nu) = \begin{cases} y^T\nu - \frac{1}{2}\|\nu\|_2^2 & \text{if } \|\frac{1}{\lambda}X^T\nu\|_\infty \le 1 \\ +\infty & \text{otherwise} \end{cases}$$

Hence the dual problem is :

$$\max_\nu y^T\nu - \frac{1}{2}\|\nu\|_2^2$$
$$\text{s.t. } \|\frac{1}{\lambda}X^T\nu\|_\infty \le 1$$

Let's consider: $Q = \frac{1}{2}I_n$ and $p = -y$
Hence $\max_\nu y^T\nu - \frac{1}{2}\|\nu\|_2^2 = \min_\nu \nu^T Q\nu + p^T\nu$
For the constraint:

$$\begin{aligned}\|\frac{1}{\lambda}X^T\nu\|_\infty \le 1 &\text{ if } \forall i \in [1,n], -1 \le \left[\frac{1}{\lambda}X^T\nu\right]_i \le 1 \\ &\text{ if } \forall i \in [1,n], \left[\frac{1}{\lambda}X^T\nu\right]_i \le 1 \text{ and } \left[-\frac{1}{\lambda}X^T\nu\right]_i \le 1 \\ &\text{ if } A\nu \preceq \lambda\mathbf{1_{2d}} \text{ with } A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}\end{aligned}$$

Finally the dual problem of LASSO is

$$\boxed{\begin{aligned}&\min_\nu \nu^T Q\nu + p^T\nu \\ &\text{s.t. } A\nu \preceq b\end{aligned}}$$

2.Implement the barrier method to solve QP.

2.1 Write a function `v_seq = centering_step(Q, p, A, b, t, v0, eps)` which implements the Newton method to solve the centering step given the inputs (Q, p, A, b), the barrier method parameter t (see lectures), initial variables v_0 and a target precision ϵ. The function outputs the sequence of variables iterates $(v_i)_{i=1,\dots,n_\epsilon}$, where n_ϵ is the number of iterations to obtain the ϵ precision. Use a backtracking line search with appropriate parameters.

2.2 Write a function `v_seq = barr_method(Q, p, A, b, v0, eps)` which implements the barrier method to solve QP using precedent function given the data inputs (Q, p, A, b), a feasible point v_0, a precision criterion ϵ. The function outputs the sequence of variables iterates $(v_i)_{i=1,\dots,n_\epsilon}$, where n_ϵ is the number of iterations to obtain the ϵ precision.

Answer:
The goal is to solve the central path problem for this unconstrained problem:

$$\min_\nu \nu^T Q\nu + p^T\nu$$
$$\text{s.t. } A\nu \preceq b$$

Hence this way:

$$\min_{\nu} g_t(\nu) = t\left(\nu^T Q\nu + p^T\nu\right) - \sum_{i=1}^{2d} \log\left(b_i - [A\nu]_i\right)$$

We consider a_i^T the rows of the matrix A

In order to implement the centering step function and the barrier method we compute the gradient and the Hessian to express them with the needed parameters : Let's compute the gradient and the Hessian matrix of the objective function:

$$\nabla g_t(\nu) = t\left(2Q\nu + p\right) + \sum_{i=1}^{2d} \left(b_i - a_i^T\nu\right)^{-1} a_i$$

we consider $\mu = \left(\left(b_1 - a_1^T\nu\right)^{-1}, \dots, \left(b_{2d} - a_{2d}^T\nu\right)^{-1}\right)^T$ we have

$$\boxed{\nabla g_t = t\left(2Q\nu + p\right) + A^T\mu}$$

For the Hessian ,

$$\nabla^2 g_t(\nu) = 2t \cdot Q + \sum_{i=1}^{2d} \left(b_i - a_i^T\nu\right)^{-2} a_i a_i^T$$

Finally:

$$\boxed{\nabla^2 g_t = 2t \cdot Q + A^T Diag\left(\mu\right)^2 A}$$