main

November 10, 2022

1 Convex Optimization - Homework 3

Report plots, comments and theoretical results in a pdf file. Send your code together with the requested functions and a main script reproducing all your experiments. You can use Matlab, Python or Julia.

Given $x_1, ..., x_n \in \mathbb{R}^d$ data vectors and $y_1, ..., y_n \in \mathbb{R}$ 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)

$$\text{minimize} \quad \tfrac{1}{2}\left\|Xw - y\right\|_2^2 + \lambda \left\|w\right\|_1 \tag{LASSO}$$

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

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

$$\begin{array}{ll} \text{minimize} & v^T Q v + p^T v \\ \text{subject to} & Av \preceq b \end{array} \tag{QP}$$

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

We can repose the LASSO problem as the following problem

$$\begin{array}{ll} \min & \tfrac{1}{2}\left\|z\right\|_2^2 + \lambda \left\|w\right\|_1 \\ \text{subject to} & z = Xw - y \end{array} \tag{LASSO'}$$

The associated Lagragian is

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

We can then calculate the dual function

$$
\begin{aligned}
g(v) &= \inf_{w,z} \frac{1}{2} z^T z + \lambda \|w\|_1 + v^T(y - Xw + z) \\
&= v^T y + \inf_z \{\frac{1}{2} z^T z + v^T z\} + \lambda \inf_w \{\|w\|_1 - \frac{1}{\lambda} v^T X w\} \\
&= \begin{cases} v^T y - \frac{1}{2} v^T v & \text{if } \left\|X^T v\right\|_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{cases}
\end{aligned}
$$

We can obtain the dual problem

$$
\text{maximize} \quad g(v) \tag{LASSO*}
$$

which can be simplified as follows

$$
\begin{array}{ll}
\min\limits_{v \in \mathbb{R}^n} & \frac{1}{2} v^T v - y^T v \\
\text{subject to} & -X^T v \leq \lambda \cdot \mathbb{1} \\
& X^T v \leq \lambda \cdot \mathbb{1}
\end{array} \tag{LASSO*}
$$

1. Impliment the barrier method to solve QP.
   - Write a function `v_seq = centering_step(Q, p, A, b, t, v0, eps)` which impliments the Newton method to solve the centering step given the inputs (Q, p, A, b), the barrier method parameter t (see lectures), initial variable v_0 and a target precision ϵ. The function outputs the sequence of variables iterates $(v_i)_{i=1,\ldots,n_\epsilon}$, where n_ϵ is the number of iterations to obtain the ϵ precision. Use a backtracking line search with appropriate parameters.
   - 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 ouptuts the sequence of variables iterates $(v_i)_{i=1,\ldots,n_\epsilon}$, where n_ϵ is the number of iterations to obtain the ϵ precision.
   - Test your function on randomly generated matrices X and observations y withz $\lambda = 10$. Plot precision criterion and gap $f(v_t) - f^*$ in semilog scale (using the best value found for f as a surrogate for f^*). Repeat for different values of the barrier method parameter $\mu = 2, 15, 50, 100, ...$ and check the impact on w. What would be an appropriate choice for μ ?