Computer Lab: Lasso

Olivier Fercoq
olivier.fercoq@telecom-paristech.fr
30 November 2017

This computer lab is in two parts.

1. Before 30 November, you should answer questions 1.1 to 2.7. This first part is to be done at home. You may work in groups as large as you wish and share your results between groups.

2. On 30 November, we will have the actual computer lab and the rest of the subject will be revealed. We booked room C45 of Telecom ParisTech. For this second part, you will be in groups of 1 or 2 persons. By 30 November at 8pm, you should:

   - send to `olivier.fercoq@telecom-paristech.fr` your code in a python file or a jupyter notebook
   - give us a paper or pdf file with the answers to the theoretical questions.

1 Data

Our goal is to solve the following Lasso problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^m} \frac{1}{2}\|Ax-b\|_2^2 \\ &\text{st: } \|x\|_1 \le r \end{aligned} \tag{1}$$

where A is the feature matrix, b is the label vector, x is the vector of parameters and $r \geq 0$ is a user-defined parameter.

Question 1.1
We will be using the dataset triazines for this computer lab. Please download the dataset on `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/regression.html#triazines`. You may wish to use `from sklearn.datasets import load_svmlight_file` to load the data.

Question 1.2
Standardize the data so that the columns of A and b are of mean 0 and standard deviation 1.

2 Proximal operator of the indicator of the ℓ_1 ball

We would like to solve the problem using the proximal gradient method. For this, we will need to compute the gradient of the loss function and the projection onto the ℓ_1 ball of radius r.

Given a vector $w \in \mathbb{R}^m$, we want to find a vector $P(w)$ solution to

$$\begin{aligned} P(w) = &\arg\min_{x \in \mathbb{R}^m} \frac{1}{2}\|x-w\|_2^2 \\ &\text{st: } \|x\|_1 \leq r \end{aligned}$$

Define the set $\Sigma_r = \{x \ : \ \sum_{i=1}^m x_i = r, x_j \geq 0, \forall j\}$, which is called the simplex. The first step is to show that one can reduce the projection on the ℓ_1 ball to the projection onto the simplex.

Question 2.1
Suppose that $\|w\|_1 \leq r$. Show that $P(w) = w$.

Question 2.2
Suppose that $\|w\|_1 > r$. Show that for all i, $(P(w))_i)$ has the same sign as w_i and that $\|P(w)\|_1 = r$. You may use a proof by contradiction.

Question 2.3
Deduce that if $\|w\|_1 > r$, then $P(w) = \text{sign}(w)\tilde{P}(|w|)$ where $\tilde{P}(|w|)$ is the projection of the vector with coordinates $|w_i|$ onto Σ_r.

We now just need to find an algorithm for the projection onto the simplex Σ_r.

$$\begin{aligned} \tilde{P}(w) = &\arg\min_{x \in \mathbb{R}^m} \frac{1}{2}\|x-w\|_2^2 \\ &\text{st: } \sum_{i=1}^m x_i = r \\ &\qquad x_j \geq 0, \forall j \end{aligned}$$

Question 2.4
Show that $x^* = \tilde{P}(w)$ if and only if there exists $\theta^* \in \mathbb{R}$ such that (x^*, θ^*) is a saddle point of the Lagrangian function

$$L(x, \theta) = \frac{1}{2}\|x-w\|_2^2 + \theta(\sum_{i=1}^m x_i - r) + \sum_{i=1}^m \iota_{\mathbb{R}_+}(x_i)\ .$$

Question 2.5
Deduce that there exists θ^* such that for all $i \in \{1, \ldots, m\}$, $x_i^* = \max(0, w_i - \theta^*)$ and $\sum_{j=1}^m \max(0, w_j - \theta^*) = r$.

Question 2.6
Show that if $w_i \geq w_j$, then $\Big((\tilde{P}(w))_i = 0 \Rightarrow (\tilde{P}(w))_j = 0\Big)$.

Question 2.7
Design an algorithm that computes $\tilde{P}(w)$. You may first sort w and work with the vector $w^\uparrow$ such that $w_1^\uparrow \leq w_2^\uparrow \leq \ldots \leq w_m^\uparrow$.

Question 2.8
Code the algorithm for the projection onto the ℓ_1 ball and run it with $w = b$ and $r = \|b\|/2$.

3 Resolution of the Lasso problem

Question 3.1
Recall the formula of the gradient of the function $x \mapsto \frac{1}{2}\|Ax - b\|_2^2$.

Question 3.2
Code the proximal gradient algorithm for Problem (1) and run it using $r = 1$. Discuss your choice of stopping criterion.

4 Comparison of two Lasso formulations

An alternative formulation of the Lasso problem is

$$\min_{x \in \mathbb{R}^m} \frac{1}{2}\|Ax - b\|_2^2 + \alpha\|x\|_1 \tag{2}$$

with $\alpha \geq 0$ a user-defined parameter. We will call this problem the penalized Lasso problem.

Question 4.1
Show that if x^* is a solution to Problem (1) with parameter r, then there exists $\alpha(r) \geq 0$ such that x^* is solution to Problem (2) with parameter $\alpha(r)$.

Conversely, show that if x^* is a solution to Problem (2) with parameter α, then there exists $r(\alpha) \geq 0$ such that x^* is solution to Problem (1) with parameter $r(\alpha)$.

Question 4.2
Code the proximal gradient algorithm for Problem (2) and run it with $\alpha = \|A^\top b\|_\infty/2$.

Question 4.3
We now split our data into 70 % for computing the parameter vector x (training set) and 30 % for the setting of the hyperparameter, i.e. r or α (validation set).

Code the holdout method for both Lasso formulations and compare the results in terms of mean square error of the final estimate and computing time.

5 Extensions

Question 5.1
What other method could be considered for the resolution of the Lasso problems (1) and (2)?