STAT 538 Homework 6

©Marina Meilă
mmp@stat.washington.edu

Problem 1 – 1-norm minimization

Consider the linear regression

$$y = X\beta + \epsilon \tag{1}$$

where $y \in \mathbb{R}^n$, $\beta \in \mathbb{R}^p$, $X \in \mathbb{R}^{n\times p}$, and $p > n$ (i.e. the problem has more unknowns $\beta_{1:p}$ than observations $y_{1:n}$. Assume X is a full rank matrix. The theory of *Compressed sensing* (see `dsp.rice.edu/cs` or future lectures) has shown that β can be estimated on the condition that it is sparse, i.e. that most of its entries are 0. However, here we will not be concerned with the theory, but only with the methods to solve for β.

a. A QP formulation for CS This optimization problem is one of the algorithms used to solve for β above.

$$(\mathcal{P}) \ \min_\beta ||\beta||_1 \quad \text{s.t.} \, ||X\beta - y||^2 \leq \delta \tag{2}$$

Apply the transformation $z = [\,\beta_+ \, \beta_- \,]^T$ and show that $(\mathcal{P})$ can be transformed into a convex optimization problem in standard form with linear objective and quadratic constraints, denoted by $(\mathcal{Q})$. How many unknowns and how many constraints has $(\mathcal{Q})$?

b. Obtain the Lagrangean of $(\mathcal{Q})$. Denote the dual variable(s) by λ.

c. In the following steps you will find the dual function and dual problem of $(\mathcal{Q})$. First, take the gradient of L w.r.t the variable z and equate to 0.

[**d. Optional - requires some linear algebra experience**] Solve for $z(\lambda)$. Be careful as the linear system you are solving does not have unique solution. Obtain the dual $g(\lambda)$ by replacing $z(\lambda)$ in the expression of L. Bring the expression to a simple form that does not depend on which $z(\lambda)$ you use.

[**e Optional.**] Obtain now the dual of $(\mathcal{Q})$, denoted by $(\mathcal{DQ})$. Show that it is concave. How many unknowns and how many constraints has $(\mathcal{DQ})$?

f. An LP formulation for CS Consider now the optimization problem

$$(\mathcal{P}') \ \min_\beta ||\beta||_1 \quad \text{s.t.} \, ||X\beta - y||_\infty \leq \delta \tag{3}$$

Apply the transformation $z = [\,\beta_+ \, \beta_- \,]^T$ and show that this can be transformed into a linear program (LP), denoted by $(\mathcal{L})$. Do not bring to standard form, i.e. do not introduce slack variables. How many unknowns and how many constraints has $(\mathcal{L})$?

g. Now bring $(\mathcal{L})$ to standard form $(\mathcal{LS})$. How many unknowns and how many constraints has $(\mathcal{LS})$?

h. Back to $(\mathcal{L})$. Obtain the Lagrangean, and take the gradient w.r.t z.

i. Now you will obtain the dual function g: The domain of g is the set of values of the dual variables for which $inf_z L(z, dual\,variables) > -\infty$. Find this domain. Find the expression of g on this domain.

j. Finally, write the dual of $(\mathcal{L})$, denoted $(\mathcal{DL})$. Make sure to include all the constraints. Reparametrize the dual so as to reduce the number of unknowns as much as possible. Bring it to a LP in standard form.

Problem 2 – Robust Least Square. B&V 4.5

Show only one equivalence of the three possible.

Problem 3 – The Markov chain as maximum entropy model

Requires working familiarity with the Markov Chain, Markov Random Fields

Let $X_1,\, X_2, \ldots X_n \in \{0,1\}$ be a collection of binary random variables. We want to define a joint distribution P over $X_{1:n}$ which has prescribed marginals, and maximum entropy.

2.1 Denote $x = (x_{1:n}) \in \{0,1\}^n$ a configuration of X, by $p_x = P(X_{1:n} = x)$ the value of the joint distribution at x, and by p the vector $p = (p_x)_{x\in\{0,1\}^n}$ The above problem can be formulated as

$$
\begin{aligned}
\max_p \quad & H(p) && (4)\\
\text{s.t.} \quad & E[X_iX_{i+1}] \;=\; r_i,\;\; i = 1:n-1 && (5)\\
& E[X_1] \;=\; r_0 && (6)\\
& E[1] \;=\; 1 && (7)
\end{aligned}
$$

Write the problem explicitly as a function of p and show that this problem is/can be formulated as a convex program.

2.2 Denote by θ_i, θ_0, θ' the Lagrange multipliers associated respectively with (5), (6), (7). Write the Lagrangean function $L(p, \theta_i, \theta_0, \theta')$ for this optimization problem, then calculate the value $p(\theta)$ that achieves $\inf_p L(p, \theta)$, for $\theta = (\theta_i, \theta_0, \theta')$.

Feel free to leave any normalization constant unevaluated (e.g replace it with $Z(\theta)$).

2.3 Examine the form of the maximum entropy solution $p(\theta)$ obtained in 2.2. Show that it represents a Markov chain.

2.4 What would happen to the solution if the constraint (6) was missing? Same question if some of the (5) were missing.

2.5 [Optional] For the graph G below, construct a Maximum Entropy problem whose solution is a MRF over $A, \ldots F$ that factors according to G. Assume all variables are binary.