Stat 375 Inference in Graphical Models

Homework 5

Due - 5/14/2012

Please return this homework in class or in the bin in Packard second floor, or by e-mail to Yiqun Liu (liuyiqun1124@gmail.com).

As mentioned in class, Gaussian BP allows to compute the minimum of a quadratic function

$$\widehat{x} = \arg\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\langle x, Qx\rangle + \langle b, x\rangle \right\}. \tag{1}$$

for $Q \in \mathbb{R}^{n\times n}$ positive definite. In this homework we will consider a case in which Q is not positive definite, but has full rank. in this case we can still define

$$\widehat{x} = -Q^{-1}b\,. \tag{2}$$

which is a stationary point (a saddle point) of the above quadratic function. The BP update equations are exactly the same as for the minimization problem. We claim that, when BP converges, it still computes the correct solution $\widehat{x}$.

We consider a specific model. An unknown signal $s_0 \in \mathbb{R}^n$ is observed in Gaussian noise

$$y = As_0 + w_0\,. \tag{3}$$

Here $y \in \mathbb{R}^m$ is a vector of observations, $A \in \mathbb{R}^{m\times n}$ is a measurement matrix, and $w_0 \in \mathbb{R}^m$ is a vector of Gaussian noise, with i.i.d. entries $w_{0,i} \sim \mathsf{N}(0, \sigma^2)$. We are given y and A, and would like to reconstruct the unknown vector s_0, and hence w_0.

A popular method consists in solving the following quadratic programming problem (known as *ridge regression*):

$$\widehat{s} = \arg\min_{s\in\mathbb{R}^n} \left\{ \frac{1}{2}\big\|y - As\big\|_2^2 + \frac{1}{2}\lambda\|s\|_2^2 \right\}. \tag{4}$$

We will do something equivalent. For $x \in \mathbb{R}^{m+n}$, $x = (z, s)$, $z \in \mathbb{R}^m$, $s \in \mathbb{R}^n$, we define

$$\mathcal{C}_{A,y}(x = (z, s)) \;=\; -\frac{1}{2}\|z\|_2^2 + \frac{1}{2}\lambda\|s\|_2^2 + \langle z, y - As\rangle\,. \tag{5}$$

We will look for the stationary point of $\mathcal{C}_{A,y}$.

(1) Show that the cost function $\mathcal{C}_{A,y}(x)$ can be written in the form

$$\mathcal{C}_{A,y}(x) = \frac{1}{2}\langle x, Qx\rangle + \langle b, x\rangle\,. \tag{6}$$

Write explicitly the form of the matrix $Q \in \mathbb{R}^{(m+n)\times(m+n)}$ and the vector $b \in \mathbb{R}^{m+n}$.

(2) Let $\widehat{x} = (\widehat{z}, \widehat{s})$ be the stationary point of $\mathcal{C}_{A,y}(z, s)$. Assuming it is unique, show that $\widehat{s}$ does coincide with the ridge estimator (4).

(3) Write a BP algorithm to compute the stationary point $\widehat{x} = (\widehat{z}, \widehat{s})$ of $\mathcal{C}_{A,y}(x)$.

(4) Generate a random matrix A, with $n = 1000$, $m = 800$ as follows. Rows are independent. Each row has $\ell = 20$ non-zero entries chosen uniformly at random. Each non-zero entry has $A_{i,j} \in \{+1, -1\}$ uniformly at random. Generate s_0 with i.i.d. standard normal distribution. Consider $\sigma^2 \in \{2, 4, 6, \dots, 100\}$ and generate random observations y.

Use the above BP algorithm to reconstruct s_0 (with $\lambda = 1/\sigma^2$). Run it for a fixed number $t = 50$ of iterations, and plot the mean square error $\|\widehat{s} - s_0\|_2^2/n$ as a function of σ^2. Denoting by $\widehat{s}^{(t)}$ the estimate after t iterations, plot the difference $\|\widehat{s}^{(t)} - \widehat{s}^{(t+1)}\|_2^2/n$ as a function of σ^2, for $t = 50$.

(5), Optional. Prove the above claim that, if BP converges, then it computes $\widehat{x}$, cf. Eq. (2) even if Q is not positive definite.