ST758, Homework 3

Due Tuesday, Sep 30, 2014

1. Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ be a positive semidefinite matrix. Show the following results.

   (a) $|a_{ij}| \leq (a_{ii} + a_{jj})/2$.

   (b) $|a_{ij}| \leq \sqrt{a_{ii} a_{jj}}$.

   (c) $\max_{i,j} |a_{ij}| = \max_i a_{ii}$.

   (d) If $a_{ii} = 0$, then $a_{ij} = a_{ji} = 0$ for all j.

   We used result (c) for the Cholesky decomposition algorithm with symmetric pivoting.

2. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the Moore-Penrose inverse $\boldsymbol{A}^+$ is the unique $n \times m$ matrix that satisfies properties $\boldsymbol{A}^+ \boldsymbol{A} \boldsymbol{A}^+ = \boldsymbol{A}^+$, $\boldsymbol{A} \boldsymbol{A}^+ \boldsymbol{A} = \boldsymbol{A}$, $(\boldsymbol{A} \boldsymbol{A}^+)^T = \boldsymbol{A} \boldsymbol{A}^+$ and $(\boldsymbol{A}^+ \boldsymbol{A})^T = \boldsymbol{A}^+ \boldsymbol{A}$.

   (a) If $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, then

   $$\boldsymbol{A}^+ = (\boldsymbol{A}^T \boldsymbol{A})^+ \boldsymbol{A}^T.$$

   (b) If the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full column rank, then

   $$(\boldsymbol{A} \boldsymbol{A}^T)^+ = \boldsymbol{A} (\boldsymbol{A}^T \boldsymbol{A})^{-2} \boldsymbol{A}^T.$$

   (c) Courrier's algorithm for computing the Moore-Penrose inverse of a matrix $\boldsymbol{M} \in \mathbb{R}^{m \times n}$ ($n \leq m$) exploits the Cholesky decomposition $\boldsymbol{L}$ of the symmetric matrix $\boldsymbol{M}^T \boldsymbol{M}$. If $\boldsymbol{M}$ has rank r, then Cholesky decomposition with symmetric pivoting yields $\boldsymbol{M}^T \boldsymbol{M} = \boldsymbol{L} \boldsymbol{L}^T$, where $\boldsymbol{L} \in \mathbb{R}^{n \times r}$ is lower triangular. Show that

   $$\boldsymbol{M}^+ = \boldsymbol{L} (\boldsymbol{L}^T \boldsymbol{L})^{-2} \boldsymbol{L}^T \boldsymbol{M}^T.$$

   (d) Write an R function implementing the Courrier's algorithm. Test your function on simulated data. Compare the speed of your function to the `ginv()` function in the `MASS` package in R, especially on rank deficient matrices.

3. (a) Read in the 'longley.dat' with the response (number of people employed) in the first column and six explanatory variables in the other columns (GNP implicit price deflator, Gross National Product, number of unemployed, number of people in the armed forces, 'noninstitutionalized' population $\geq$ 14 years of age, year). Include an intercept in your model.

   (b) Assuming linear model $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$, compute the regression coefficients $\hat{\boldsymbol{\beta}}$, their standard errors, and variance estimate $\hat{\sigma}^2$ using following methods: QR decomposition, Cholesky decomposition, and sweep operator. Please compute them directly using numerical linear algebra functions; you can use the "black-box" function `lm()` only to check your results. (Hint: `chol2inv()` function computes the inverse of a matrix from its Cholesky factor.)

(c) Find out which method is the `lm()` function in R using? And which algorithm is being used?

(d) One popular regularization method is the ridge regression, which estimates regression coefficients by minimizing a penalized least squares criterion

$$\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{\beta}\|_2^2.$$

Show that the ridge solution is given by

$$\hat{\boldsymbol{\beta}}_\lambda = (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{y}.$$

(e) Show that the ridge estimator is equivalent to the solution of a regular least squares problem with added observations. Compute the ridge regression estimates $\hat{\boldsymbol{\beta}}_\lambda$ at two different values of λ by solving this augmented least squares problem. You can use any method of your choice (QR, Cholesky, or sweep).