ST758, Homework 2

Due Oct 15, 2013

1. (a) Read in the file 'longley.dat' on course webpage, which has the response $\boldsymbol{y}$ in the first column and six explanatory variables in the other columns.
   (b) Compute the 6×6 sample covariance matrix and call it $\boldsymbol{V}$.
   (c) Compute the 6×6 correlation coefficient matrix $\boldsymbol{C}$ from $\boldsymbol{V}$. What do you observe in $\boldsymbol{C}$?
   (d) Partition the matrix $\boldsymbol{V}$ as
   $$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix},$$
   where the blocks $\boldsymbol{V}_{ij}$ have size 3×3. Compute $\boldsymbol{V}_{22} - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{V}_{12}$, using Cholesky decomposition.
   (e) Compute $\boldsymbol{V}_{22} - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{V}_{12}$ again, using sweeping.
   (f) Assume linear model $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$. Include an intercept in your model and compute the regression coefficients $\hat{\boldsymbol{\beta}}$, their standard errors, variance estimate $\hat{\sigma}^2$, fitted values $\hat{\boldsymbol{y}}$, and residuals $\hat{\boldsymbol{e}}$ using three methods – Cholesky, QR, and sweeping. Please compute them directly; you can use other "black-box" function, e.g. `lm()`, only to check.
   (g) Let $\boldsymbol{X}_i \in \mathbb{R}^{n \times i}$ contain the first i columns of the design matrix. That is $\boldsymbol{X}_1 = \boldsymbol{1}_n$, $\boldsymbol{X}_2 = (\boldsymbol{1}_n, \boldsymbol{x}_1)$, $\boldsymbol{X}_3 = (\boldsymbol{1}_n, \boldsymbol{x}_1, \boldsymbol{x}_2)$, and so on. Compute $\boldsymbol{y}^T \boldsymbol{P}_{\boldsymbol{X}_1} \boldsymbol{y}$, $\boldsymbol{y}^T(\boldsymbol{P}_{\boldsymbol{X}_i} - \boldsymbol{P}_{\boldsymbol{X}_{i-1}})\boldsymbol{y}$, $i = 2, \ldots, 7$, and $\boldsymbol{y}^T(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}$, where $\boldsymbol{P}_{\boldsymbol{A}}$ denotes the orthogonal projection onto the column space of a matrix $\boldsymbol{A}$. (Hint: don't ever think about forming these projection matrices.)

2. Let $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, $n \geq p$, be the design matrix in linear regression. The least squares solution is given by
   $$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X\beta}\|_2^2.$$
   (a) Show that $\hat{\boldsymbol{\beta}}$ is a least squares solution if and only if it satisfies the normal equation $\boldsymbol{X}^\intercal \boldsymbol{X} \hat{\boldsymbol{\beta}} = \boldsymbol{X}^\intercal \boldsymbol{y}$.
   (b) If $\boldsymbol{X}$ has rank $r < p$, the least squares solution is not unique. Show that any vector $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal \boldsymbol{X})^- \boldsymbol{X}^\intercal \boldsymbol{y}$, where $(\boldsymbol{X}^\intercal \boldsymbol{X})^-$ is any generalized inverse, is a least squares solution, and the residual vector $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is invariant to the choice of the generalized inverse.
   (c) Assume $\boldsymbol{X}$ has rank $r < p$ and the *QR decomposition with (column) pivoting* yields
   $$\boldsymbol{X} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0}_{(n-r)\times r} & \boldsymbol{0}_{(n-r)\times(p-r)} \end{pmatrix} \boldsymbol{\Pi}^\intercal, \tag{1}$$
   where $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is orthogonal, $\boldsymbol{R}_{11} \in \mathbb{R}^{r \times r}$ is upper triangular with positive diagonal entries, and $\boldsymbol{\Pi} \in \mathbb{R}^{p \times p}$ is a permutation matrix. Show that
   $$\boldsymbol{X}^- = \boldsymbol{\Pi} \begin{pmatrix} \boldsymbol{R}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{Q}^\intercal$$

is a generalized inverse of $\boldsymbol{X}$. Is this generalized inverse the Moore-Penrose inverse of $\boldsymbol{X}$? The `qr()` function in R performs QR decomposition with (column) pivoting. Study the documentation of `qr()` and do the following on the Longley data in Q1.

(d) Add an extra column (sum of the intercept and the first predictor) to the original design matrix such that $\boldsymbol{X}_{\text{new}} = (\mathbf{1}_n, \boldsymbol{x}_1, \mathbf{1}_n + \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_6)$. Compute *a* least squares solution using the generalized inverse in (c) obtained from QR. Compare the least squares solution, fitted values, and residuals to those obtained in Q1(f).

(e) Given the singular value decomposition (SVD)

$$\boldsymbol{X} = \boldsymbol{U} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0}_{r\times(p-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(p-r)} \end{pmatrix} \boldsymbol{V}^{\mathsf{T}},$$

where $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{U} \in \mathbb{R}^{p\times p}$ are orthogonal, and $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{r\times r}$ is diagonal with positive diagonal entries. Show that

$$\boldsymbol{X}^{+} = \boldsymbol{V} \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(p-r)\times r} & \mathbf{0}_{(p-r)\times(n-r)} \end{pmatrix} \boldsymbol{U}^{\mathsf{T}}$$

is the Moore-Penrose inverse of $\boldsymbol{X}$.

(f) Show that $\hat{\boldsymbol{\beta}} = \boldsymbol{X}^{+}\boldsymbol{y}$ is a least squares solution and has the minimum ℓ_2 norm among all least squares solutions.

(g) Redo part (d) but using the Moore-Penrose inverse obtained from SVD. Compare the least squares solution, fitted values, and residuals to those obtained in Q2(d).

3. Let $\boldsymbol{X} \in \mathbb{R}^n$ be a random vector with i.i.d. standard normal entries.

(a) Show that $\boldsymbol{X}$ has density

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}/2}.$$

(b) Any affine transformation $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{\mu}$ of $\boldsymbol{X}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{\mu} \in \mathbb{R}^m$, is called a multivariate normal random vector. Assume $\text{rank}(\boldsymbol{A}) = m$. Show that

$$\mathbf{E}(\boldsymbol{Y}) = \boldsymbol{\mu}, \quad \mathbf{Var}(\boldsymbol{Y}) = \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} = \boldsymbol{\Omega},$$

and the density of $\boldsymbol{Y}$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \left(\frac{1}{2\pi}\right)^{m/2} |\det(\boldsymbol{\Omega})|^{-1/2} e^{-(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})/2}.$$

(c) Suppose $\boldsymbol{Y}$ is partitioned as two sub-vectors $\boldsymbol{Y} = (\boldsymbol{Y}_1^{\mathsf{T}}, \boldsymbol{Y}_2^{\mathsf{T}})^{\mathsf{T}}$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^{\mathsf{T}}, \boldsymbol{\mu}_2^{\mathsf{T}})^{\mathsf{T}}$ and $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}$ are partitioned compatibly. Show that the conditional distribution of $\boldsymbol{Y}_2$ given $\boldsymbol{Y}_1$ is normal with mean and variance

$$\begin{aligned} \mathbf{E}(\boldsymbol{Y}_2 \mid \boldsymbol{Y}_1) &= \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2 \\ \mathbf{Var}(\boldsymbol{Y}_2 \mid \boldsymbol{Y}_1) &= \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}. \end{aligned}$$

Let $\boldsymbol{Y} \sim N(\mathbf{0}_n, \boldsymbol{\Omega}_n)$, where $\boldsymbol{\Omega}_n = (r^{|i-j|})_{i,j}$. Use $m = 10$ and $r = 0.9$ in below.

(d) Generate a $\boldsymbol{y}$ from this model, using only the `rnorm()` function.

(e) Partition $\boldsymbol{y}$ as $(\boldsymbol{y}_1^T, \boldsymbol{y}_2^T)^T$, where $\boldsymbol{y}_1 \in \mathbb{R}^6$ and $\boldsymbol{y}_2 \in \mathbb{R}^4$. Compute the following quantities at the generated $\boldsymbol{y}$, using either Cholesky or sweeping.

- Marginal density $f_{\boldsymbol{Y}_1}(\boldsymbol{y}_1)$
- Conditional mean $\mathbf{E}(\boldsymbol{Y}_2 \mid \boldsymbol{Y}_1)$ and conditional variance $\mathbf{Var}(\boldsymbol{Y}_2 \mid \boldsymbol{Y}_1)$
- Joint density $f_{\boldsymbol{Y}}(\boldsymbol{y})$
- Conditional density $f_{\boldsymbol{Y}_2|\boldsymbol{Y}_1}(\boldsymbol{y}_2|\boldsymbol{y}_1)$
- Let $\boldsymbol{Z} = \boldsymbol{Y}_2 - \mathbf{E}(\boldsymbol{Y}_2|\boldsymbol{Y}_1)$. Evaluate density $f_{\boldsymbol{Z}}(\boldsymbol{z})$ at $\boldsymbol{z} = \boldsymbol{y}_2 - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{y}_1$

(f) Now you have gained some experience using sweeping or Cholesky to handle multivariate normal computations. Re-think part (e). Are there more efficient ways to compute the multivariate normal density for this covariance structure? If you find one, re-evaluate the joint density $f_{\boldsymbol{Y}}(\boldsymbol{y})$ using your method.