STA302/STA1001, Week 8

Mark Ebden, 26 October 2017 - morning

With grateful acknowledgment to Alison Gibbs

This week's content

- Midterms
  - We won't discuss these until both tests have occurred
  - Posting to Portal the test paper .pdf, and solutions, can be expected on Friday 27 October
- Entering Chapter 5
  - Exercise to recap what we know so far about matrices
  - Pages 26–28 of the RMA (*Review of Matrix Algebra*) .pdf file
  - Matrix SLR

Exercise

Recall from last week: What is $\mathrm{var}(\mathbf{AX})$?

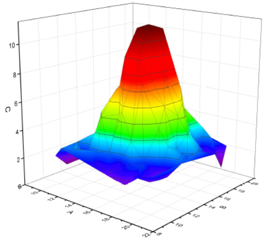
This week's content

- Midterms
  - We won't discuss these until both tests have occurred
  - Posting to Portal the test paper .pdf, and solutions, can be expected on Friday 27 October
- Entering Chapter 5
  - Exercise to recap what we know so far about matrices
  - **Pages 26–28 of the RMA (*Review of Matrix Algebra*) .pdf file**
  - Matrix SLR

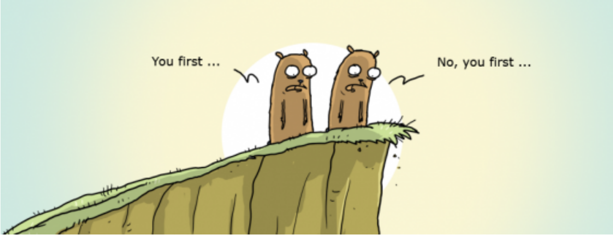
Matrix differentiation

If $\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$ and $f(\boldsymbol{\theta})$ is a scalar, then

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_k} \end{pmatrix}$$



Two gradient lemmas



Lemma 1: Suppose $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$ and $f(\boldsymbol{\theta}) = \mathbf{c}'\boldsymbol{\theta} = \sum_{i=1}^{n} c_i\theta_i$. Then,

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{c}$$

Lemma 2: Suppose $\mathbf{A}$ is a $k \times k$ symmetric matrix, and $f(\boldsymbol{\theta}) = \boldsymbol{\theta}'\mathbf{A} boldsymbol \theta$. Then,

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}$$

Using Matrix SLR

Recall our question from last week: How do we solve the least-squares estimates of the regression coefficients, in matrix form?

In other words: in *matrix form* we seek b_0 and b_1 that minimize the sum of squares of residuals,

$$\text{RSS} = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Least-squares estimation of regression coefficients

Let's start by speaking of RSS in terms of $\boldsymbol{\beta}$, and getting rid of the summation, with:

$$\mathrm{RSS}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Why this works: Easiest is to begin with the matrix RHS, multiply it out, and arrive at the $\sum$ LHS. Building on last week:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Note that there can arise notation collisions when dealing with matrices and random variables simultaneously. There is no universally accepted way around this.

Least-squares estimation of regression coefficients

Let's continue:

$$\begin{aligned}
\mathrm{RSS}(\beta) &= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \\
&= (\mathbf{Y}' - \beta'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\beta) \quad \text{from pp 22 \& 24 of RMA} \\
&= \mathbf{Y}'\mathbf{Y} - \beta'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \\
&= \mathbf{Y}'\mathbf{Y} - 2\beta'\mathbf{X}'\mathbf{Y} + \beta'\mathbf{X}'\mathbf{X}\beta \\
\frac{\partial \mathrm{RSS}(\beta)}{\partial \beta} &= 0 - 2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta}
\end{aligned}$$

Setting the derivative to zero as before,

$$\begin{aligned}
2\mathbf{X}'\mathbf{X}\hat{\beta} &= 2\mathbf{X}'\mathbf{Y} \qquad . \\
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}
\end{aligned}$$

major drawback: (X'X) has to have an inverse, or it has to be invertible, full rank

A reflection on the inverse

Should we always set $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$?

See Answer Slides 1-2.

A closer look at $\mathbf{X}'\mathbf{X}$

What does $\mathbf{X}'\mathbf{X}$ simplify to? Is it symmetric?

Recall: $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$

See Answer Slide 3.

A closer look at $(\mathbf{X}'\mathbf{X})^{-1}$

What does $(\mathbf{X}'\mathbf{X})^{-1}$ simplify to? Recall:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \Longrightarrow \qquad A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

See Answer Slide 4.

Bringing this together

Given our expression for $(\mathbf{X}'\mathbf{X})^{-1}$, what is $\hat{\boldsymbol{\beta}}$ and hence our $\hat{\beta}_0$ and $\hat{\beta}_1$?

Recall that

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Next steps

- HW2 (not for credit) will be posted on Portal in the weekend of 27-29 October
- Next week we'll continue in Chapter 5, covering all of it eventually
- Reminder: no TA office hours on 27/30 October, and Portal contains midterm-return information