STK3100 Exercises, Week 1

Vinnie Ko, Jonas Moss

September 7, 2018

These solutions are based on the solutions Vinnie Ko wrote for the 2017 iteration of the GLM course.

Exercise 1.7

Recall the definition of *model space*:

$$C(\boldsymbol{X}) = \{\boldsymbol{\eta} : \text{there is a } \boldsymbol{\beta} \text{ such that } \boldsymbol{\eta} = \boldsymbol{X\beta}\}.$$

No matter how our $\boldsymbol{X}$ looks like, we can always obtain $\boldsymbol{\eta} = \boldsymbol{X\beta}$ by letting $\boldsymbol{\beta} = \boldsymbol{0}$. So, $\boldsymbol{0}$ is in the model space $C(\boldsymbol{X})$ for any linear model $\boldsymbol{\mu} = \boldsymbol{X\beta}$.

Exercise 1.8

Let $\boldsymbol{X}$ be an arbitrary model matrix with dimension $n \times p$ and let $\boldsymbol{x}_{*p}$ denote the last column of $\boldsymbol{X}$. Let $\boldsymbol{X}_0$ be $\boldsymbol{X}$ without $\boldsymbol{x}_{*p}$. The dimension of $\boldsymbol{X}_0$ is then $n \times (p-1)$.
Now, let an arbitrary $\boldsymbol{b} \in C(\boldsymbol{X}_0)$. Then, by definition, there exists $\boldsymbol{\delta}_0 \in \mathbb{R}^{(p-1)}$ such that $\boldsymbol{b} = \boldsymbol{X}_0\boldsymbol{\delta}_0$.
Let $\boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}_0 \\ 0 \end{bmatrix} \in \mathbb{R}^p$, then $\boldsymbol{X\delta} = \begin{bmatrix} \boldsymbol{X}_0 & \boldsymbol{x}_{*p} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\delta}_0 \\ 0 \end{bmatrix} = \boldsymbol{X}_0\boldsymbol{\delta}_0 = \boldsymbol{b}$.
So, every $\boldsymbol{b} \in C(\boldsymbol{X}_0)$ satisfies $\boldsymbol{b} \in C(\boldsymbol{X})$. Therefore, $C(\boldsymbol{X}_0) \subset C(\boldsymbol{X})$.

Exercise 1.11

The model matrix

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{1}_{n_1} & \boldsymbol{1}_{n_1} & \boldsymbol{0}_{n_1} & \cdots & \boldsymbol{0}_{n_1} \\ \boldsymbol{1}_{n_2} & \boldsymbol{0}_{n_2} & \boldsymbol{1}_{n_2} & \cdots & \boldsymbol{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{1}_{n_{r-1}} & \boldsymbol{0}_{n_{r-1}} & \boldsymbol{0}_{n_{r-1}} & \cdots & \boldsymbol{1}_{n_{r-1}} \\ \boldsymbol{1}_{n_r} & -\boldsymbol{1}_{n_r} & -\boldsymbol{1}_{n_r} & \cdots & -\boldsymbol{1}_{n_r} \end{bmatrix}$$

has linearly independent columns (i.e. has full rank). So, $\boldsymbol{\beta}$ is identifiable.
(The notation $\boldsymbol{0}_{n_r}$ and $\boldsymbol{1}_{n_r}$ are as defined in section 1.3.3 of the book.)

Exercise 1.12

(a)

In this *two-way layout* setting, we have 2 categorical variables, one with r categories and one with c categories. These 2 variables are dummy coded by β_i and γ_j respectively. So the parameter vector is $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_r, \gamma_1, \cdots, \gamma_c]^{\mathrm{T}}$. So, $p = r + c + 1$.

Since the model is *balanced*, $\boldsymbol{y}$ and the model matrix look like

$$\boldsymbol{Y} = \begin{bmatrix} Y_{1,1,1} \\ \vdots \\ Y_{1,1,n} \\ Y_{1,2,1} \\ \vdots \\ Y_{1,2,n} \\ \vdots \\ Y_{1,r,1} \\ \vdots \\ Y_{1,r,n} \\ Y_{2,1,1} \\ \vdots \\ Y_{r,c,n} \end{bmatrix}, \quad \boldsymbol{X} = \left[\begin{array}{c|cccc|cccc} \mathbf{1}_n & \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n & \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{1}_n & \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n & \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_n & \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n \\ \hline \mathbf{1}_n & \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n & \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{1}_n & \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n & \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_n & \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n & \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n & \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n \end{array}\right]$$

and it has rank $r + c - 1 < p$. (The sum of columns 2 to $r + 1$ equals the first column and so does the sum of columns r+2 to r+c+1.) So, $\boldsymbol{\beta}$ is not identifiable.

Exercise 1.13

(a)

$$\boldsymbol{X} = \left[\begin{array}{c|c|cc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{array}\right], \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_2 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}$$

Interpretation:
β_0: $\mathrm{E}[Y|A = 1, B = 1]$.
β_i: $\mathrm{E}[Y|A = i] - \mathrm{E}[Y|A = 1]$.
γ_j: $\mathrm{E}[Y|B = j] - \mathrm{E}[Y|B = 1]$.

(b)

$$
\boldsymbol{X} = \left[\begin{array}{c|c|cc}
1 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 \\
1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
\hline
1 & -1 & 1 & 0 \\
1 & -1 & 1 & 0 \\
1 & -1 & 0 & 1 \\
1 & -1 & 0 & 1 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1
\end{array}\right], \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \gamma_1 \\ \gamma_2 \end{bmatrix}, \quad \beta_2 = -\beta_1, \quad \gamma_3 = -\gamma_1 - \gamma_2
$$

Interpretation:
β_0: $\mathrm{E}\left[Y\right]$.
β_i: $\mathrm{E}\left[Y|A = i\right] - \mathrm{E}\left[Y\right]$.
γ_j: $\mathrm{E}\left[Y|B = j\right] - \mathrm{E}\left[Y\right]$.

(c)

In (a) and (b), $\boldsymbol{X}$ has full-rank with $\mathrm{rank}(\boldsymbol{X}) = r + c - 1 = 4$.

Additional Exercise I

Jonas' perspective

This is a small note about this exercise is about differentiating matrices and vectors. The concept introduced in this note is not part of the curriculum, but should be of interest anyway.

The exercise is about finding the derivative of a function with respect to a vector. But what does that mean? In order to make sense of this, we must answer the question:

What is this kind of derivative?

Let $f : U \to Y$ be a map between to normed spaces X and Y, where $U \subseteq X$ is open. A bounded linear map $A_x : X \to Y$ satisfying

$$
\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - A_x h\|}{\|h\|}
$$

is a *Frchet derivative* at x. This definition understands the derivative at a *point* x as a linear map that approximates the function arbitrarily well if its argument is close enough to zero.

Fact: This is the *natural definition* of a derivative for normed spaces, including $\mathbb{R}^n$, $\mathcal{H}$, and $\mathbb{C}$. A holomorphic function is a Frchet differentiable function from $\mathbb{C} \supseteq U \to \mathbb{C}$, while the Jacobian at x is the Frchet derivative of a function $f : U \supseteq \mathbb{R}^n \to \mathbb{R}^m$.

You can read more about this kind of derivative in Tom Lindstrm's excellent lecture notes.

(https://www.uio.no/studier/emner/matnat/math/MAT2400/v16/spaces.pdf)

In our exercise, $X = \mathbb{R}^n$ and $Y = \mathbb{R}$.

The exercise

Part I Here $f : \mathbb{R}^n \to \mathbb{R}$ is the linear functional a, with operation $x \mapsto x^T a$. By the equation above we'll need to look at $\frac{\|f(x+h) - f(x) - A_x h\|}{\|h\|}$. Now $f(x+h) - f(x) - A_x h$ equals $x^T a + h^T a - x^T a - A_x h$. This equals 0 when $A_x h = h^T a$, hence a is the derivative, and is independent of x.

Part II Now $f : \mathbb{R}^n \to \mathbb{R}$ is the quadratic functional $x \mapsto x^T Bx$ for some matrix B. In this case, $f(x+h) - f(x) - A_x h$ equals

$$x^T Bx + h^T Bh + x^T Bh + h^T Bx - x^T Bx - Ax$$

Here $x^T Bx$ cancels, and since $x^T Bh = h^T B^T x$, we obtain $h^T Bh + h^T Bx + h^T B^T x$. Since $\left\|h^T Bh\right\| = O\left(h^2\right)$, $Ah = h^T\left(B + B^T\right)x$ is a natural guess. You can verify that this is the derivative by the triangle inequality.

Vinnie's solution

This solution is more in line with the statement of the exercise. Use it if you don't care about theoretical mathematics.

It's given that

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,p} \\ \vdots & \ddots & \vdots \\ a_{p,1} & \cdots & a_{p,p} \end{bmatrix}$$

and

$$\frac{\partial \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \dfrac{\partial \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}}{\partial \beta_p} \end{bmatrix}, \quad \frac{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \dfrac{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta}}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta}}{\partial \beta_p} \end{bmatrix}.$$

i)

Since

$$\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta} = \sum_{j=1}^{p} a_j \beta_j,$$

we have

$$\frac{\partial \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \dfrac{\partial \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}}{\partial \beta_p} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \sum_{j=1}^{p} a_j \beta_j}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \sum_{j=1}^{p} a_j \beta_j}{\partial \beta_p} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = \mathbf{a}.$$

ii)

First, we obtain the element-wise expression for the right side term.

$$(\mathbf{A}+\mathbf{A}^{\mathrm{T}})\boldsymbol{\beta} = \mathbf{A}\boldsymbol{\beta} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\beta} = \begin{bmatrix} \sum_{j=1}^{p} a_{1,j}\beta_j \\ \vdots \\ \sum_{j=1}^{p} a_{p,j}\beta_j \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{p} a_{i,1}\beta_i \\ \vdots \\ \sum_{i=1}^{p} a_{i,p}\beta_i \end{bmatrix}.$$

Then, we show that the left side term is equal to it. Since

$$\boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta} = \begin{bmatrix}\beta_1 \cdots \beta_p\end{bmatrix} \cdot \begin{bmatrix} a_{1,1} & \cdots & a_{1,p} \\ \vdots & \ddots & \vdots \\ a_{p,1} & \cdots & a_{p,p} \end{bmatrix} \cdot \begin{bmatrix}\beta_1 \\ \vdots \\ \beta_p\end{bmatrix} = \sum_{i=1}^{p} \beta_i \sum_{j=1}^{p} a_{i,j}\beta_j = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{i,j}\beta_i\beta_j,$$

we have

$$\begin{aligned}
\frac{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} &= \begin{bmatrix} \frac{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta}}{\partial \beta_1} \\ \vdots \\ \frac{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta}}{\partial \beta_p} \end{bmatrix} = \begin{bmatrix} \frac{\partial \sum_{i=1}^{p}\sum_{j=1}^{p} a_{i,j}\beta_i\beta_j}{\partial \beta_1} \\ \vdots \\ \frac{\partial \sum_{i=1}^{p}\sum_{j=1}^{p} a_{i,j}\beta_i\beta_j}{\partial \beta_p} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial a_{i,j}\beta_i\beta_j}{\partial \beta_1} \\ \vdots \\ \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial a_{i,j}\beta_i\beta_j}{\partial \beta_p} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{p} a_{1,j}\beta_j + \sum_{i=1}^{p} a_{i,1}\beta_i \\ \vdots \\ \sum_{j=1}^{p} a_{p,j}\beta_j + \sum_{i=1}^{p} a_{i,p}\beta_i \end{bmatrix} \\
&= (\mathbf{A}+\mathbf{A}^{\mathrm{T}})\boldsymbol{\beta}.
\end{aligned}$$

For further career in statistics, it is handy to know the following matrix differentiation rules:

Let the scalar α be defined by $\alpha = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{b}$ and $\boldsymbol{A}$ are not a function of $\mathbf{x}$, then

$$\frac{\partial \boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}, \quad \frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{b}}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{b}, \quad \frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}.$$

These three rules are actually a special case of a more general rule:

Let the scalar α be defined by $\alpha = \mathbf{u}^{\mathrm{T}}\boldsymbol{A}\mathbf{v}$ where
$\mathbf{u} = \mathbf{u}(\boldsymbol{x}), \mathbf{v} = \mathbf{v}(\boldsymbol{x})$ and $\mathbf{u} : \mathbb{R}^m \to \mathbb{R}^m, \mathbf{v} : \mathbb{R}^n \to \mathbb{R}^n$, then

$$\frac{\partial \mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{v}}{\partial \boldsymbol{x}} = \frac{\partial \mathbf{u}}{\partial \boldsymbol{x}}\mathbf{A}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \boldsymbol{x}}\mathbf{A}^{\mathrm{T}}\mathbf{u}.$$

Note that there are several conventions in matrix calculus. In this solution, we stick to the *denominator* layout (a.k.a. *Hessian* formulation).