Notes 5

Maryclare Griffin

3/2/2023

These notes are based on Chapters 1, 5, and 6 of KNNL.

Recall from the previous notes, the linear regression model for a dependent variable or response Y and independent variables, predictors, or covariates X_1, … X_{p-1} is defined as:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i$$

where:

- $\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters
- $X_{i1}, \dots, X_{i,p-1}$ are known constants
- ϵ_i is a random error term with mean $E\{\epsilon_i\} = 0$ and variance $\sigma^2\{\epsilon_i\} = \sigma^2$; ϵ_i and ϵ_j are uncorrelated so that their covariance is zero (i.e., $\sigma\{\epsilon_i, \epsilon_j\} = 0$ for all i, j; $i \neq j$)
- $i = 1, \dots, n$

Remember, we don't observe $\beta_0, \beta_1, \dots, \beta_{p-1}$ in the real world. Instead, we **estimate** them by finding the values $b_0, b_1, \dots, b_{p-1}$ that minimize the sum of squared deviations of the response values Y_i from the regression function $\beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik}$ with respect to $\beta_0, \beta_1, \dots, \beta_{p-1}$.

In the previous set of notes, we showed how the `lm` function in `R` can be used to compute $b_0, b_1, \dots, b_{p-1}$. We also showed how closed form equations can be derived for b_0 and b_1 when $p = 2$. What about when we have multiple predictors? We want to find the values $b_0, b_1, \dots, b_{p-1}$ that solve all p of the **normal equations**:

- $\sum_{i=1}^n -2Y_i + 2b_0 + 2\left(\sum_{k=1}^{p-1} b_k X_{ik}\right) = 0$
- For $k > 0$, $\sum_{i=1}^n -2Y_i X_{ik} + 2b_k X_{ik}^2 + 2b_0 X_{ik} + 2\sum_{l=1, l\neq k}^{p-1} b_l X_{il} X_{ik} = 0$

First, let's rewrite the equations a bit, getting rid of the extra twos and starting with the terms involving elements of $\boldsymbol{b}$ and/or $\boldsymbol{X}$:

- $\sum_{i=1}^n b_0 + \sum_{k=1}^{p-1} b_k X_{ik} - Y_i = 0$
- For $k > 0$, $\sum_{i=1}^n b_k X_{ik}^2 + b_0 X_{ik} + \sum_{l=1, l\neq k}^{p-1} b_l X_{il} X_{ik} - 2Y_i X_{ik} = 0$

For this, we'll need linear algebra. Linear algebra lets us write out sums and sets of equations efficiently. We will define:

- The $n \times 1$ column vector $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$
- The $n \times p$ **design matrix** $\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{pmatrix}$

- The $p \times 1$ vector $\boldsymbol{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{pmatrix}$

  Note: Matrix multiplication gives us a nice way of computing several sums at once. If $\boldsymbol{A}$ is an $r \times c$ matrix and $\boldsymbol{B}$ is a $c \times s$ matrix, the product $\boldsymbol{D} = \boldsymbol{AB}$ is an $r \times s$ matrix with elements:

$$d_{ij} = \sum_{k=1}^{c} a_{ik} b_{kj}.$$

This looks a bit familiar, and suggests that we can write out our normal equations using matrix multiplication. The normal equations involve sums $\sum_{i=1}^{n} Y_i X_{ik}$ for different values of k, as well as sums $\sum_{i=1}^{n} b_k X_{ik}^2$, and $\sum_{i=1}^{n} \sum_{l=1, l \neq k}^{p-1} b_l X_{il} X_{ik}$. The first term resembles what we would expect to obtain if we multiplied $\boldsymbol{X}$ and $\boldsymbol{Y}$, however their dimensions are not amenable to matrix multiplication as-is.

Note: The transpose of a vector or matrix is obtained by interchanging the rows and columns. * The transpose of the $n \times 1$ column vector $\boldsymbol{Y}$ is the $1 \times n$ row vector ormatrix $\boldsymbol{Y}' = \begin{pmatrix} Y_1 & Y_2 & \dots & Y_n \end{pmatrix}$ * The transpose of the $n \times p$ matrix $\boldsymbol{X}$ is $\boldsymbol{X}' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1,p-1} & X_{2,p-1} & \dots & X_{n,p-1} \end{pmatrix}$

We can compute $\boldsymbol{X}'\boldsymbol{Y}$, let's try it!

$$\boldsymbol{X}'\boldsymbol{Y} = \begin{pmatrix} \sum_{i=1}^{n} Y_k \\ \sum_{i=1}^{n} Y_k X_{i1} \\ \vdots \\ \sum_{i=1}^{n} Y_k X_{i,p-1} \end{pmatrix}$$

Aha! We have the $\sum_{i=1}^{n} Y_k$ that appears in the first normal equation and the $\sum_{i=1}^{n} Y_k X_{ik}$ terms that appear in the remaining $p-1$ normal equations. What about the terms involving $\boldsymbol{b}$ and elements of $\boldsymbol{X}$? A natural quantity to consider is $\boldsymbol{Xb}$. Let's compute that!

$$\boldsymbol{Xb} = \begin{pmatrix} b_0 + \sum_{k=1}^{p-1} b_k X_{1k} \\ b_0 + \sum_{k=1}^{p-1} b_k X_{2k} \\ \vdots \\ b_0 + \sum_{k=1}^{p-1} b_k X_{nk} \end{pmatrix}$$

This isn't quite what we're looking for. We are missing terms involving squares X_{ik}^2 and products $X_{ik} X_{il}$ for

$l \neq k$. This suggests seeing what we get if we multiply by $\boldsymbol{X}'$ to get a $p \times 1$ vector:

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1,p-1} & X_{2,p-1} & \dots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} b_0 + \sum_{k=1}^{p-1} b_k X_{1k} \\ b_0 + \sum_{k=1}^{p-1} b_k X_{2k} \\ \vdots \\ b_0 + \sum_{k=1}^{p-1} b_k X_{nk} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n} b_0 + \sum_{k=1}^{p-1} b_k X_{ik} \\ \sum_{i=1}^{n} X_{i1} \left(b_0 + \sum_{k=1}^{p-1} b_k X_{ik} \right) \\ \sum_{i=1}^{n} X_{i2} \left(b_0 + \sum_{k=1}^{p-1} b_k X_{ik} \right) \\ \vdots \\ \sum_{i=1}^{n} X_{i,p-1} \left(b_0 + \sum_{k=1}^{p-1} b_k X_{ik} \right) \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n} b_0 + \sum_{k=1}^{p-1} b_k X_{ik} \\ \sum_{i=1}^{n} b_1 X_{i1}^2 + b_0 X_{i1} + \sum_{l=2}^{p-1} b_l X_{i1} X_{il} \\ \sum_{i=1}^{n} b_2 X_{i2}^2 + b_0 X_{i2} + b_1 X_{i1} X_{i2} + \sum_{l=3}^{p-1} b_l X_{ik} X_{il} \\ \vdots \\ \sum_{i=1}^{n} b_{p-1} X_{i,p-1}^2 + b_0 X_{i,p-1} + \sum_{l=1}^{p-2} b_l X_{il} X_{i,p-1} \end{pmatrix}
\end{aligned}$$

The first element of $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}$ minus the first element of $\boldsymbol{X}'\boldsymbol{Y}$ set equal to 0 gives us the first normal equation! The second element of $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}$ minus the second element of $\boldsymbol{X}'\boldsymbol{Y}$ set equal to 0 gives us the first normal equation! And so on!

Note: Two vectors or matrices are said to be equal if they have the same dimension and all of the corresponding elements are equal, i.e. if $\boldsymbol{A}$ is an $r \times c$ matrix and $\boldsymbol{B}$ is an $r \times c$ matrix, then elements of $\boldsymbol{A} = \boldsymbol{B}$ indicates that $a_{ij} = b_{ij}$ for $i = 1, \dots r$ and $j = 1, \dots c$.

Note: The difference of two vectors or matrices of the same dimensions is the difference of their elements, i.e. if $\boldsymbol{A}$ is an $r \times c$ matrix and $\boldsymbol{B}$ is an $r \times c$ matrix, then elements of $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B}$ satisfy $d_{ij} = a_{ij} - b_{ij}$.

Accordingly, if $\mathbf{0}$ is a $p \times 1$ vector with all elements exactly equal to zero, we can write all p of the normal equations simultanously using linear algebra as:

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} - \boldsymbol{X}'\boldsymbol{Y} = \mathbf{0}.$$

This equation is a "nice'' function of $\boldsymbol{b}$! A few more linear algebra facts will allow us to solve it.

Note: The sum of two vectors or matrices of the same dimensions is the sum of their elements, i.e. if $\boldsymbol{A}$ is an $r \times c$ matrix and $\boldsymbol{B}$ is an $r \times c$ matrix, then elements of $\boldsymbol{D} = \boldsymbol{A} + \boldsymbol{B}$ satisfy $d_{ij} = a_{ij} + b_{ij}$.

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{Y}$$

Note: The **identity matrix** or **unit matrix** is denoted by $\boldsymbol{I}$. It is a diagonal matrix whose elements on the main diagonal I_{kk} are all equal to 1, and remaining elements are equal to 0. Premultiplying or postmultiplying any $r \times r$ matrix $\boldsymbol{A}$ by the $r \times r$ identity matrix $\boldsymbol{I}$ leaves $\boldsymbol{A}$ unchanged, i.e. $\boldsymbol{I}\boldsymbol{A} = \boldsymbol{A}$ and $\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}$.

Note: The inverse of a $r \times r$ square matrix $\boldsymbol{A}$ is another $r \times r$ square matrix, denoted by $\boldsymbol{A}^{-1}$, such that $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$. The inverse matrix $\boldsymbol{A}^{-1}$ exists if the matrix $\boldsymbol{A}$ is rank r, i.e. if $\boldsymbol{A}$ is **nonsingular** or **full rank**. A matrix with rank less than r is said to be **singular** or **not of full rank**. The inverse of a matrix with with rank r also has rank r.

If $\boldsymbol{X}'\boldsymbol{X}$ is full rank, then we have a closed form solution for $\boldsymbol{b}$:

$$\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

Note: $\boldsymbol{X}'\boldsymbol{X}$ is full rank when there we cannot write any column of $\boldsymbol{X}$ as a linear combination of the remaining columns of $\boldsymbol{X}$. Practically, this means that $\boldsymbol{X}'\boldsymbol{X}$ is never full rank when $n < p$.

Example 1: Let's return to the data from portrait studios in 21 cities run by Dwaine Studios, Inc. The studios specialize in portraits of children. Let X_1 be the number of persons aged 16 or younger in a city, let X_2 refer to per capita disposable income in a city, and let Y be the sales of portraits of children in that city from one of the 21 studies. We're going to construct an design matrix and compute b_0, b_1, and b_2 by hand.

```
load("~/Dropbox/Teaching/STAT525/Spring2023/bookdata/dwaine.RData")
X1 <- data$X1 # Extract the first predictor
X2 <- data$X2 # Extract the second predictor
Y <- data$Y # Extract the response
n <- length(Y) # Record the number of observations
X <- cbind(rep(1, n), X1, X2)
XtY <- t(X)%*%Y # Compute X'Y
XtX <- t(X)%*%X # Compute X'X
XtX.inv <- solve(XtX) # Invert XtX
b <- XtX.inv%*%XtY # Solve for b
b0 <- b[1] # Extract the intercept
b1 <- b[2] # Extract b1
b2 <- b[3] # Extract b2
```

We obtain estimates $b_0 = -68.857$, $b_1 = 1.455$, and $b_2 = 9.366$. These are the same numbers we got from `lm` previously!

Now that we've introduced linear algebra, it's worth noting that we can write the linear regression model and Q, the sum of squares criterion using linear algebra.

Letting $\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$, the linear regression model for a dependent variable or response Y and independent variables, predictors, or covariates $X_1, \dots X_{p-1}$ is defined as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where:

- The elements of $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$ are parameters
- The elements of the $n \times p$ matrix $\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{pmatrix}$ are known constants

- For $i = 1, \ldots, n$, ϵ_i is a random error term with mean $E\{\epsilon_i\} = 0$ and variance $\sigma^2\{\epsilon_i\} = \sigma^2$; ϵ_i and ϵ_j are uncorrelated so that their covariance is zero (i.e., $\sigma\{\epsilon_i, \epsilon_j\} = 0$ for all i, j; $i \neq j$)

The $p \times 1$ vector of estimated regression coefficients $\boldsymbol{b}$ corresponds to the value of $\boldsymbol{\beta}$ that minimizes

$$Q = (\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$$

Accordingly, the p normal equations are obtained by taking the derivative of $\boldsymbol{Q}$ with respect to $\boldsymbol{\beta}$ and setting them equal to $\mathbf{0}$. We can see this by first expanding out Q, and then taking derivatives.

$$Q = \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X}\boldsymbol{\beta} - (\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{Y} + (\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{X}\boldsymbol{\beta})$$

Note: Given two $1 \times p$ vectors $\boldsymbol{a}$ and $\boldsymbol{c}$, $\boldsymbol{a}'\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{a}$.

$$Q = \boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{X}\boldsymbol{\beta})$$

Note: The transpose of a product of matrices $(\boldsymbol{AB})'$ is obtained by reversing the order of the elements in the product and taking the transpose of each, $\boldsymbol{B}'\boldsymbol{A}'$.

$$Q = \boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{Y}'\boldsymbol{X}\beta + \beta'\boldsymbol{X}'\boldsymbol{X}\beta$$

Note: Given an $1 \times p$ vector $\boldsymbol{A}$, the derivative of $\boldsymbol{A\beta}$ with respect to $\boldsymbol{\beta}$ is $\boldsymbol{A}$.

$$\frac{\partial Q}{\partial \beta} = -2\boldsymbol{Y}'\boldsymbol{X} + \frac{\partial Q}{\partial \beta}\beta'\boldsymbol{X}'\boldsymbol{X}\beta$$

Note: Given an $1 \times p$ vector $\boldsymbol{A}$, the derivative of $\boldsymbol{\beta}'\boldsymbol{A\beta}$ with respect to $\boldsymbol{\beta}$ is $2\boldsymbol{\beta}'\boldsymbol{A}$.

$$\frac{\partial Q}{\partial \beta} = -2\boldsymbol{Y}'\boldsymbol{X} + 2\beta'\boldsymbol{X}'\boldsymbol{X}$$

If we substitute $\boldsymbol{b}$ in for $\boldsymbol{\beta}$, this looks very similar to the expression we got for the normal equations but transposed and multiplied by 2. Substituting $\boldsymbol{b}$ in for $\boldsymbol{\beta}$ and taking the transpose, we get:

$$-2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \mathbf{0}$$

Dividing by 2 and rearranging yields our normal equations:

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} - \boldsymbol{X}'\boldsymbol{Y} = \mathbf{0}$$

Ta-da!