Go over HW3
- coding categorical variables
- model.matrix

Lecture 9: Least Squares Estimation using Matrices
STAT 632, Spring 2020

Mathematical Preliminaries

Let

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}$$

be an $n \times 1$ vector, and $f(\boldsymbol{\theta})$ a scalar function of $\boldsymbol{\theta}$. The derivative of $f(\boldsymbol{\theta})$ with respect to the vector $\boldsymbol{\theta}$ is defined as

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_n} \end{pmatrix}$$

Recall from calculus two important rules of differentiation for a scalar a and variable x:

1. $\frac{d}{dx} ax = a$
2. $\frac{d}{dx} ax^2 = 2ax$

There are similar rules when taking a derivative with respect to a vector:

1. Let $\boldsymbol{c}' = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix}$ and $f(\boldsymbol{\theta}) = \boldsymbol{c}'\boldsymbol{\theta}$, then it follows that

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

2. Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix and $f(\boldsymbol{\theta}) = \boldsymbol{\theta}'\boldsymbol{A}\boldsymbol{\theta}$, then it follows that

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

Your turn: Prove the first rule for vector differentiation.

$$f(\theta) = c'\theta = c_1\theta_1 + c_2\theta_2 + \dots + c_n\theta_n$$

$$\frac{\partial f(\theta)}{\partial \theta} = \begin{pmatrix} \partial f(\theta)/\partial\theta_1 \\ \partial f(\theta)/\partial\theta_2 \\ \vdots \\ \partial f(\theta)/\partial\theta_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \underset{n\times 1}{c}$$

Recall the matrix notation for multiple linear regression (MLR):

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

Your turn: What are the dimensions of each term?

$Y \Rightarrow n \times 1$ $\quad$ $\beta = (p+1) \times 1$

$X \Rightarrow n \times (p+1)$ $\quad$ $e \Rightarrow n \times 1$

For MLR, the residuals sum of squares, as a function of the vector $\boldsymbol{\beta}$, can be written in matrix notation as

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

Expanding this equation out gives:

$$\begin{aligned} R(\boldsymbol{\beta}) &= \boldsymbol{Y'Y} - \boldsymbol{Y'X\beta} - (\boldsymbol{X\beta})'\boldsymbol{Y} + (\boldsymbol{X\beta})'(\boldsymbol{X\beta}) \\ &= \boldsymbol{Y'Y} - \boldsymbol{Y'X\beta} - \boldsymbol{\beta'X'Y} + \boldsymbol{\beta'X'X\beta} \\ &= \underset{①}{\boldsymbol{Y'Y}} - \underset{②}{2\boldsymbol{Y'X\beta}} + \underset{③}{\boldsymbol{\beta'X'X\beta}} \end{aligned}$$

since $(AB)' = B'A'$

Your turn: Argue that $\boldsymbol{Y'X\beta} = \boldsymbol{\beta'X'Y}$, so the terms can be combined.

$$\underset{1 \times n}{Y'} \; \underset{n \times (p+1)}{X} \; \underset{(p+1) \times 1}{\beta}$$

$\Rightarrow 1 \times 1$ scalar

$$Y'X\beta = (Y'X\beta)' = (X\beta)'Y = \beta'X'Y$$

Next, to find the least squares estimates, we minimize $R(\boldsymbol{\beta})$ by taking the derivative, with respect to the vector $\boldsymbol{\beta}$, and setting the derivative equal to zero. Since $\boldsymbol{X'X}$ is a symmetric matrix, we can use the two rules for vector differentiation to accomplish this.

$$\frac{\partial R(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{0} - 2\boldsymbol{X'Y} + 2\boldsymbol{X'X\beta}$$

$$-2\underbrace{\underset{(p+1) \times n}{X'} \; \underset{n \times 1}{Y}}_{(p+1) \times 1} + 2\underbrace{\underset{(p+1) \times n}{X'} \; \underset{n \times (p+1)}{X} \; \underset{(p+1) \times 1}{\beta}}_{(p+1) \times 1}$$

Setting the derivative equal to zero gives the **normal equations** for MLR:

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

Assuming that $\boldsymbol{X'X}$ is invertible, the vector of least squares estimates is given by

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

① $\dfrac{\partial Y'Y}{\partial \beta} = 0$

since derivative of a constant is 0

② $\dfrac{\partial Y'X\beta}{\partial \beta} = X'Y$

by first rule:

$\dfrac{\partial c'\theta}{\partial \theta} = c$ with $c' = \underset{1 \times n}{Y'} \underset{n \times (p+1)}{X}$

$\theta = \beta$

③ $\dfrac{\partial \beta' X' X \beta}{\partial \beta} = 2X'X\beta$

by second rule:

$\dfrac{\partial \theta' A \theta}{\partial \theta} = 2A\theta$

$A = X'X$

$\theta = \beta$