Matrix Derivatives:

Week 1

$f: \mathbb{R}^p \to \mathbb{R} \qquad f(x) = a^T x$

$\nabla f(x) = a$

$g: \mathbb{R}^p \to \mathbb{R} \qquad g(x) = x^T A x$, where $A_{p\times p}$ is symmetric

$\nabla g(x) = 2Ax$

$g''(x) = 2A$

Ex:

$$z = \| y_{n\times 1} - X_{n\times p}\, \beta_{p\times 1} \|^2 = y^T y - 2y^T X\beta + \beta^T X^T X \beta$$

$$\frac{\partial z}{\partial \beta} = -2X^T y + 2X^T X\beta = 0$$

$$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T y$$

$$\frac{\partial^2 z}{\partial \beta\, \partial \beta^T} = 2X^T X$$

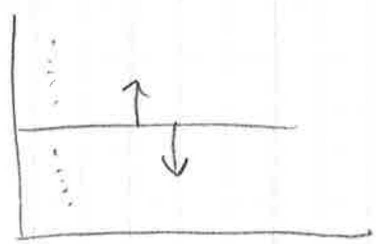
$$a^T X^T X a = \| Xa \|^2 \geq 0 \quad \text{(positive definite)}$$

$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T y$ is a minimum

Centering by matrix multiplication

$$\underset{n\times 1}{J_n} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

y



$$\min \| y - J_n \beta \|^2 \Rightarrow \hat{\beta} = (J_n^T J_n)^{-1} J_n^T y = \bar{y}$$

$$y - J_n \bar{y} = y - J_n (J_n^T J_n)^{-1} J_n^T y = (I - J_n (J_n^T J_n)^{-1} J_n^T) y$$

$$\underbrace{\frac{1}{n} J_n^T}_{\text{gives average}} (I - J_n (J_n^T J_n)^{-1} J_n^T) y = \frac{1}{n} (J_n^T - \cancel{J_n^T J_n (J_n^T J_n)^{-1}} J_n^T) y$$

$$= \frac{1}{n} (J_n^T - J_n^T) y = 0$$

premultiply any vector by $(I - J_n(J_n^T J_n)^{-1} J_n^T)$ mean centers it.

$(I - J_n(J_n^T J_n)^{-1} J_n^T)\, X_{n\times p}$ mean centers the columns of X

$X(I - J_p(J_p^T J_p)^{-1} J_p^T)$ mean centers the rows of X

$$I - J_n(\underbrace{J_n^T J_n}_{n})^{-1} J_n^T = I - \frac{1}{n} J_{n\times n}$$

Variance via matrix multiplication

$$s^2 = \frac{\Sigma(y_i - \bar{y})^2}{n-1} = \frac{\|y - \bar{y}J_n\|^2}{n-1} = \frac{1}{n-1}\tilde{y}^T\tilde{y}$$

$$\tilde{y} = (I - J_n(J_n^T J_n)^{-1} J_n^T)y$$

$$\tilde{y}^T\tilde{y} = y^T\underbrace{(I - J_n(J_n^T J_n)^{-1} J_n^T)}_{\text{Symmetric}}(I - J_n(J_n^T J_n)^{-1} J_n^T)y$$

$$= y^T(\underbrace{I - J_n(J_n^T J_n)^{-1} J_n^T}_{\text{idempotent}})y$$ = sum of squared deviations around the mean

$X_{n\times p}$

$$X^T(I - \underbrace{J_n(J_n^T J_n)^{-1} J_n^T}_{H})X = \underbrace{X^T(I-H)^T}_{\tilde{X}^T}\underbrace{(I-H)X}_{\tilde{X}} = X^T(I-H)(I-H)X$$

$$= \tilde{X}^T\tilde{X} \Rightarrow \frac{1}{n-1}\tilde{X}^T\tilde{X} = Cov(X)$$