Example: Linear algebra and derivatives

Prof. Yao Xie

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Multiple linear regression

- set-up

$$y_i = \beta_1 x_{i1} + \ldots \beta_p x_{ip} + \beta_0 + \epsilon_i, \quad i = 1, \ldots, n$$

coefficient for p variables: $\beta = [\beta_0, \beta_1, \cdots, \beta_p]^\mathsf{T}$
n samples: $(y_i, x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$

- Find coefficient using least-square method

$$\min_\beta \sum_{i=1}^{n} (y_i - (\beta_1 x_{i1} + \ldots \beta_p x_{ip} + \beta_0))^2$$

- matrix-vector form

$$y = A\beta + \epsilon, \quad A = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix} \in \mathbb{R}^{n\times(p+1)}$$

- solving optimization problem

$$\min_\beta \|y - A\beta\|_2^2$$

$$
\begin{aligned}
\|y - A\beta\|_2^2 &= (y - A\beta)^T (y - A\beta) \\
&= y^T y + \beta^T A^T A\beta - \beta^T A^T y - y^T A\beta \\
&= y^T y + \beta^T A^T A\beta - 2y^T A\beta
\end{aligned}
$$

Take the derivative with respect to β, using rules

$$
\frac{\partial x^T A}{\partial x} = A, \quad \frac{\partial x^T A x}{\partial x} = 2Ax.
$$

Then we have

$$
\begin{gathered}
\frac{\partial \|y - A\beta\|_2^2}{\partial \beta} \\
= 2A^T A\beta - 2A^T y
\end{gathered}
$$

(Please note that $y^T y$ is a constant term with regard to β)