Collection of Problems that I think are Cool

Benjamin Chu

May 12, 2019

1 Statistics

Problem 1.1

Consider a multiple regression where $n > p$ and $rank(\mathbf{X}) = p$. Let

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2$$

where $\mathbf{e} = (e_1, \ldots, e_n)^t = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ are the regression residuals and $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$. Show that $\hat{\sigma}^2$ is an unibased estimator of σ^2.

Proof. We have

$$\hat{\sigma^2} = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-p}(\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}).$$

Also, $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{H}\mathbf{y}$. Repeatedly applying cyclic permuation and linearity of trace operator, we have

$$\begin{aligned}
&\mathrm{E}\left((\mathbf{y}-\mathbf{H}\mathbf{y})^T(\mathbf{y}-\mathbf{H}\mathbf{y})\right) = \mathrm{E}(\mathbf{y}^T(\mathbf{I}-\mathbf{H})(\mathbf{I}-\mathbf{H})\mathbf{y}) = \mathrm{E}(\mathbf{y}^T(\mathbf{I}-\mathbf{H})\mathbf{y}) \\
&= \mathrm{tr}\left(\mathrm{E}(\mathbf{y}^T(\mathbf{I}-\mathbf{H})\mathbf{y})\right) = \mathrm{E}\left(\mathrm{tr}((\mathbf{X}\boldsymbol{\beta}+\varepsilon)^T(\mathbf{I}-\mathbf{H})(\mathbf{X}\boldsymbol{\beta}+\varepsilon))\right) \\
&= \mathrm{E}\left(\mathrm{tr}((\mathbf{X}\boldsymbol{\beta}+\varepsilon)^T(\mathbf{X}+\varepsilon-\mathbf{H}\mathbf{X}\boldsymbol{\beta}-\mathbf{H}\varepsilon))\right) = \mathrm{E}\left(\mathrm{tr}((\mathbf{X}\boldsymbol{\beta}+\varepsilon)^T(\mathbf{I}-\mathbf{H})\varepsilon)\right) \\
&= \mathrm{E}\left(\mathrm{tr}(\varepsilon^T(\mathbf{I}-\mathbf{H})\varepsilon)\right) = \mathrm{tr}\left((\mathbf{I}-\mathbf{H})\,\mathrm{E}(\varepsilon\varepsilon^T)\right) = \mathrm{tr}\left((\mathbf{I}-\mathbf{H})\,\mathrm{Var}(\varepsilon)\right) \\
&= \sigma^2\,\mathrm{tr}(\mathbf{I}-\mathbf{H}) = \sigma^2\left(\mathrm{tr}(\mathbf{I}_{n\times n}) - \mathrm{tr}(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\right) = \sigma^2(n-p).
\end{aligned}$$

□

Problem 1.2

Let $\mathbf{X} \in \mathbb{R}^{n\times p}$, $\lambda_i \in \mathbb{R}$, and $\mathbf{x}_i^T \in \mathbb{R}^p$ be a row of $\mathbf{X}$. Show that

$$\sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mathbf{x}_i^T = \mathbf{X}^T \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_n \end{bmatrix} \mathbf{X}$$

Problem 1.3

Suppose $f \in C^2$. Show that there exists $y \in (x_0, x)$ such that:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(y)(x - x_0)^2.$$

This exact formula for 2nd order Taylor's expansion motivates the quadratic upper bound principle, which is used ubiquitously in MM algorithms.

Proof. Applying fundamental theorem of calculus twice, we have

$$\begin{aligned} f(x) &= f(x_0) + \int_{x_0}^{x} f'(x_1)dx_1 \\ &= f(x_0) + \int_{x_0}^{x} \left(f'(x_0) + \int_{x_0}^{x_1} f''(x_2)dx_2 \right) dx_1 \\ &= f(x_0) + f'(x_0)(x - x_0) + \int_{x_0}^{x} \int_{x_0}^{x_1} f''(x_2)dx_2 dx_1. \end{aligned}$$

By mean value theorem, there exists $y \in (x_0, x_1)$ such that $\int_{x_0}^{x_1} f''(x_2)dx_2 = f''(y)(x_1 - x_0)$. Thus

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \int_{x_0}^{x} f''(y)(x_1 - x_0)dx_1 \\ &= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(y)(x - x_0). \end{aligned}$$

$\square$