Collection of Problems that I think are Cool

Benjamin Chu

May 13, 2019

1 Math

Problem 1.1

Let $\mathbf{X} = \mathbb{R}^{n\times n}$ random matrix. Show that probability that $\det(\mathbf{X}) = 0$ is 0. That is, almost all $n \times n$ random matrices are invertible.

2 Statistics

Problem 2.1

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 $\mathbf{e} = (e_1, ..., e_n)^t = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ 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

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

□

Problem 2.2

Show that sample mean and sample variance are 2 independent statistics.

3 Useful Tricks and Identities

Problem 3.1

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 3.2 Exact 2nd order Taylor's expansion

Suppose $f \in C^2(\mathbb{R})$. 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 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

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

□

Problem 3.3 Clever use of Cauchy-Schwarz [Lange, 2016, Exercise 1.4.18]

Prove the majorization

$$\begin{aligned} (x+y-z)^2 \le -(x_n+y_n-z_n)^2 + 2(x_n+y_n-z_n)(x+y-z) \\ +3[(x-x_n)^2 + (y-y_n)^2 + (z-z_n)^2] \end{aligned}$$

which separtes the variables x, y, and z. In examples 1.3.6 and 1.3.7 this would facilitate penalizing parameter curvature rather than changes in parameter values.

Proof. First, move the first two terms on the right to the left:

$$(x+y-z)^2 - 2(x_n+y_n-z_n)(x+y-z) + (x_n+y_n-z_n)^2 \le 3[(x-x_n)^2 + (y-y_n)^2 + (z-z_n)^2]$$

The left can be factored cleanly as

$$\big((x+y-z) - (x_n+y_n-z_n)\big)^2 \le 3[(x-x_n)^2 + (y-y_n)^2 + (z-z_n)^2]$$

$$\iff (a+b+c)^2 \le 3a^2 + 3b^2 + 3c^2$$

where $a = x - x_n, b = y - y_n, c = z - z_n$. Now define $v = (1,1,1), u = (a,b,c)$. By Cauchy-Schwarz, we obtain the desired result:

$$(a+b+c)^2 \le 3(a^2+b^2+c^2).$$

□

4 Real worl Application Problems

Problem 4.1

Suppose we wish to fit a large n small p linear regression problem. Every day millions of new sample points are generated. How would one obtain $\hat{\beta}$ without saving larger and larger matrices?

References

[Lange, 2016] Lange, K. (2016). *MM optimization algorithms*, volume 147. SIAM.