STATS 3001 / STATS 4101 / STATS 7054
Statistical Modelling III
Tutorial 1
2022
Solutions

QUESTIONS:

1. Suppose A and B are matrices of dimension $n \times m$ and $m \times n$ respectively. Prove that the trace satisfies

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

SOLUTIONS:

Observe first that the ith diagonal element of AB is given by

$$\sum_{j=1}^{m} a_{ij} b_{ji}$$

and hence

$$\mathrm{tr}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ji}$$

On the other hand, the jth diagonal element of BA is

$$\sum_{i=1}^{n} b_{ji} a_{ij}$$

so that

$$\mathrm{tr}(BA) = \sum_{j=1}^{m} \sum_{i=1}^{n} b_{ji} a_{ij} = \mathrm{tr}(AB)$$

as required.

---

2. If A is a constant $n \times n$ matrix and $\mathcal{Y}$ is an $n \times n$ matrix whose elements are random variables, prove that

$$E[\mathrm{tr}(A\mathcal{Y})] = \mathrm{tr}(AE[\mathcal{Y}]).$$

SOLUTIONS:

$$\begin{aligned} E[\mathrm{tr}(A\mathcal{Y})] &= E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}Y_{ji}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}E(Y_{ij}) \\ &= \mathrm{tr}(AE[\mathcal{Y}]) \end{aligned}$$

as required.

---

3. If X is a matrix of dimension $n \times p$ with linearly independent columns (so that X^TX is invertible) then prove that the matrix

$$P = X(X^TX)^{-1}X^T$$

satisfies

$$P^T = P = P^2$$

and hence show that

$$(I - P)^T = (I - P) = (I - P)^2.$$

SOLUTIONS:
Observe first that (X^TX) is symmetric so that $(X^TX)^{-1}$ is also symmetric. Then

$$P^T = (X(X^TX)^{-1}X^T)^T = (X^T)^T((X^TX)^{-1})^T(X)^T = X(X^TX)^{-1}X^T = P$$

and

$$P^2 = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T) = X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T =$$

Finally,

$$(I - P)^T = I^T - P^T = I - P$$

and

$$(I - P)^2 = I + P^2 - 2P = I + P - 2P = I - P.$$

---

4. Suppose $\boldsymbol{y} \in \mathbb{R}^n$ and X is a $n \times p$ matrix with linearly independent columns. Prove that

$$Q(\boldsymbol{\beta}) = \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2$$

is uniquely minimized by

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

(a) Algebraically by considering

$$Q(\boldsymbol{\beta}) = \|\boldsymbol{y} - X\hat{\boldsymbol{\beta}} + X\hat{\boldsymbol{\beta}} - X\boldsymbol{\beta}\|^2.$$

(a) Using calculus to set up system of equations

$$\frac{\partial Q}{\partial \beta_j} = 0, \quad j = 1, 2, \ldots, p$$

and solving for $\boldsymbol{\beta}$.

Hint: This can be done easily by observing (or proving) first that

$$\frac{\partial}{\partial \boldsymbol{u}} \boldsymbol{u}^T \boldsymbol{u} = 2\boldsymbol{u}^T$$

and then applying the chain rule of multivariable calculus to

$\boldsymbol{u} = (\boldsymbol{y} - X\boldsymbol{\beta})$ to obtain

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2(\boldsymbol{y} - X\boldsymbol{\beta})^T X.$$

SOLUTIONS:
(a)

$$\begin{aligned}
Q(\boldsymbol{\beta}) &= \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 \\
&= \|\boldsymbol{y} - X\hat{\boldsymbol{\beta}} + X\hat{\boldsymbol{\beta}} - X\boldsymbol{\beta}\|^2 \\
&= \|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\|^2 + \|X\hat{\boldsymbol{\beta}} - X\boldsymbol{\beta}\|^2 + 2(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^T (X\hat{\boldsymbol{\beta}} - X\boldsymbol{\beta})
\end{aligned}$$

Now,

$$\begin{aligned}(\boldsymbol{y}-X\hat{\boldsymbol{\beta}})^T(X\hat{\boldsymbol{\beta}}-X\boldsymbol{\beta}) &= ((I-P)\boldsymbol{y})^T(X(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})) \\ &= \boldsymbol{y}^T(I-P)X(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ &= \boldsymbol{y}^T(I-X(X^TX)^{-1}X^T)X(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ &= \boldsymbol{y}^T(X-X(X^TX)^{-1}X^TX)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ &= \boldsymbol{y}^T[0](\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ &= 0\end{aligned}$$

and hence

$$\begin{aligned}Q(\boldsymbol{\beta}) &= \|\boldsymbol{y}-X\hat{\boldsymbol{\beta}}\|^2+\|X\hat{\boldsymbol{\beta}}-X\boldsymbol{\beta}\|^2 \\ &\geq \|\boldsymbol{y}-X\hat{\boldsymbol{\beta}}\|^2\end{aligned}$$

with equality if and only if

$$\|X\hat{\boldsymbol{\beta}}-X\boldsymbol{\beta}\|^2=0.$$

That is, if and only if,

$$\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}$$

since the columns of X are linearly independent.

(b)

Following the hint, consider first

$$f(\boldsymbol{u})=\boldsymbol{u}^T\boldsymbol{u}=\sum_{i=1}^{n}u_i^2.$$

It follows that

$$\frac{\partial f}{\partial u_i}=2u_i$$

so, in vector notation

$$\frac{\partial f}{\partial \boldsymbol{u}}=2\boldsymbol{u}^T.$$

Next observe if $\boldsymbol{u}=(\boldsymbol{y}-X\boldsymbol{\beta})$ then

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\beta}} = -X$$

so

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\boldsymbol{u}^T X = -2(\boldsymbol{y} - X\boldsymbol{\beta})^T X = -2(\boldsymbol{y}^T X - \boldsymbol{\beta}^T X^T X)$$

and

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \mathbf{0} \Rightarrow \boldsymbol{y}^T X = \boldsymbol{\beta}^T X^T X \Rightarrow (X^T X)\boldsymbol{\beta} = X^T \boldsymbol{y}$$

so that the solution is

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

Note that a careful proof also requires us to check that the solution is a minimum rather than some other type of stationary point. This can be done by considering the second derivative but we have already confirmed this in (i).
