STATS 3001 / STATS 4101 / STATS 7054
Statistical Modelling III
Tutorial 6
2022
Solutions

QUESTIONS:

1. ## 0.1 Singular value decomposition

The SVD of the $N \times p$ matrix X has the form

$$X = UDV^T,$$

where the columns of U and V are orthogonal, *i.e.*

$$U^T U = I \text{ and } V^T V = I,$$

and D is a diagonal matrix with diagonal entries $d_1 \geq d_2 \geq \ldots \geq d_p \geq 0$.

(a) Show that for linear regression

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

(b) Show that for ridge regression:

$$X\hat{\boldsymbol{\beta}}_\lambda = UD(D^2 + \lambda I)^{-1} DU^T \boldsymbol{y}$$

(c) Hence, show that

$$X\hat{\boldsymbol{\beta}}_\lambda = \sum_{j=1}^{p} \boldsymbol{u}_j \frac{d_j^2}{d_j^2 + \lambda} \boldsymbol{u}_j^T \boldsymbol{y},$$

where $\boldsymbol{u}_j$ are the columns of U.

SOLUTIONS:

(a) First note that

$$X^T X = VDU^T UDV^T = VD^2 V^T$$

So we have

$$
\begin{aligned}
X(X^TX)^{-1}X^T\boldsymbol{y} &= UDV^T(VD^2V^T)^{-1}VDU^T\boldsymbol{y} \\
&= UDV^T(V^T)^{-1}D^{-2}V^{-1}VDU^T\boldsymbol{y} \\
&= UU^T\boldsymbol{y}
\end{aligned}
$$

(b)

$$
\begin{aligned}
X\hat{\boldsymbol{\beta}}_\lambda &= X(X^TX+\lambda I)^{-1}X^T\boldsymbol{y} \\
&= UDV^T(VD^2V^T+\lambda I)^{-1}(UDV^T)^T\boldsymbol{y} \\
&= UDV^T(VD^2V^T+\lambda VV^T)^{-1}VDU^T\boldsymbol{y} \\
&= UDV^T(V(D^2+\lambda I)V^T)^{-1}VDU^T\boldsymbol{y} \\
&= UDV^T(V^T)^{-1}(D^2+\lambda I)^{-1}V^{-1}VDU^T\boldsymbol{y} \\
&= UD(D^2+\lambda I))DU^T\boldsymbol{y}
\end{aligned}
$$

(c)

First note that

$$
D(D^2+\lambda I)^{-1}D = \begin{bmatrix} \frac{d_1^2}{d_1^2+\lambda} & & \\ & \ddots & \\ & & \frac{d_1^2}{d_1^2+\lambda} \end{bmatrix}
$$

Hence

$$
UD(D^2+\lambda I)^{-1}D = \begin{bmatrix} \frac{u_{11}d_1^2}{d_1^2+\lambda} & \frac{u_{12}d_2^2}{d_2^2+\lambda} & \cdots & \frac{u_{1p}d_p^2}{d_p^2+\lambda} \\ \frac{u_{21}d_1^2}{d_1^2+\lambda} & \frac{u_{22}d_2^2}{d_2^2+\lambda} & \cdots & \frac{u_{2p}d_p^2}{d_p^2+\lambda} \\ \vdots & \ddots & & \vdots \\ \frac{u_{n1}d_1^2}{d_1^2+\lambda} & \frac{u_{n2}d_2^2}{d_2^2+\lambda} & \cdots & \frac{u_{np}d_p^2}{d_p^2+\lambda} \end{bmatrix}
$$

Finally we have

$$
UD(D^2+\lambda I)^{-1}DU^T = \begin{bmatrix} \frac{u_{11}d_1^2}{d_1^2+\lambda} & \frac{u_{12}d_2^2}{d_2^2+\lambda} & \cdots & \frac{u_{1p}d_p^2}{d_p^2+\lambda} \\ \frac{u_{21}d_1^2}{d_1^2+\lambda} & \frac{u_{22}d_2^2}{d_2^2+\lambda} & \cdots & \frac{u_{2p}d_p^2}{d_p^2+\lambda} \\ \vdots & \ddots & & \vdots \\ \frac{u_{n1}d_1^2}{d_1^2+\lambda} & \frac{u_{n2}d_2^2}{d_2^2+\lambda} & \cdots & \frac{u_{np}d_p^2}{d_p^2+\lambda} \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{n1} \\ u_{12} & u_{22} & \cdots & u_{n2} \\ \vdots & & \ddots & \vdots \\ u_{1p} & u_{2p} & \cdots & u_{np} \end{bmatrix}
$$

$$
\Rightarrow [UD(D^2+\lambda I)^{-1}DU^T]_{ij} = \sum_{k=1}^{p} u_{ik}\frac{d_k^2}{d_k^2+\lambda}u_{jk}
$$

Consider the ijth element of $\boldsymbol{u}_k \frac{d_k^2}{d_k^2+\lambda} \boldsymbol{u}_k^T$:

$$\left[\boldsymbol{u}_k \frac{d_k^2}{d_k^2+\lambda} \boldsymbol{u}_k^T\right]_{ij} = \begin{bmatrix} \frac{u_{1k} d_k^2}{d_k^2+\lambda} \\ \frac{u_{2k} d_k^2}{d_k^2+\lambda} \\ \vdots \\ \frac{u_{nk} d_k^2}{d_k^2+\lambda} \end{bmatrix} \begin{bmatrix} u_{1k} & u_{2k} & \dots & u_{nk} \end{bmatrix}_{ij}$$
$$= u_{ik} \frac{d_k^2}{d_k^2+\lambda} u_{jk}$$

So we have

$$[UD(D^2+\lambda I)^{-1} DU^T]_{ij} = \sum_{k=1}^{p} \left[\boldsymbol{u}_k \frac{d_k^2}{d_k^2+\lambda} \boldsymbol{u}_k^T\right]_{ij}$$

So we have

$$UD(D^2+\lambda I)^{-1} DU^T = \sum_{k=1}^{p} \boldsymbol{u}_k \frac{d_k^2}{d_k^2+\lambda} \boldsymbol{u}_k^T$$
