3. Consider the matrix

$$X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}.$$

(a) Find the matrix $X^T X$ and diagonalize it.

(b) Find the SVD of X.

(c) Find the best approximation to X with rank equal to 1.

a)

$$X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

$$X^T = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}.$$

$$X^X X = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}.$$

$$\det(X^T X - \lambda I) = \det\begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = 0.$$

$$(6-\lambda)^2 - 1 = 0$$

$$36 - 12\lambda + \lambda^2 - 1 = 0.$$

$$\lambda^2 - 12\lambda + 35 = 0.$$

$$\lambda_{1,2} = \frac{12 \pm 2}{2}$$

$$\lambda_1 = 7$$

$$\lambda_2 = 5.$$

$$D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}.$$

$\lambda_1 = 7:$

$$(X^T X - 7 I) V = 0.$$

$$\begin{pmatrix} 6-7 & 1 \\ 1 & 6-7 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \doteq 0.$$

$$\begin{cases} -1 \cdot V_1 + 1 \cdot V_2 = 0. \\ 1 \cdot V_1 - 1 \cdot V_2 = 0. \end{cases}$$

$$V_1 = V_2.$$

$$V = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$\|V\| = \sqrt{1^2 + 1^2} = \sqrt{2}.$$

normalize: $V = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$

$\underline{\lambda_2 = 5:}$

$$(X^T X - 5 \cdot I) V = 0.$$

$$\begin{pmatrix} 6-5 & 1 \\ 1 & 6-5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

$$v_1 = -v_2.$$

$$v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\|v\| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}.$$

$$v = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

$$V = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \qquad V^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

$$X^T X = VDV^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

2. Consider the model $y = X\beta + u$ where β is non-random, $\mathbb{E}(u \mid X) = 0$, the matrix X of size $n \times k$ has rank $X = k$, but $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 W$ with $W \neq I$. Let $\hat{\beta}$ be the standard OLS estimator of β.

   (a) Find $\mathbb{E}(\hat{\beta} \mid X)$, $\mathbb{E}(\hat{\beta})$.

   (b) Find $\mathbb{V}\mathrm{ar}(\hat{\beta} \mid X)$.

   (c) How do you think, will the standard confidence interval for β be valid in this case?

   (d) Find $\mathbb{C}\mathrm{ov}(y, \hat{\beta} \mid X)$.

a) $E(\hat{\beta}|X) = E((X^T X)^{-1} X^T y \,|X) = (X^T X)^{-1} X^T E(y|X) =$

$$= (X^T X)^{-1} X^T E(X\beta + u | X) = (X^T X)^{-1} E(X\beta | X) +$$

$$+ (X^T X)^{-1} E(u|X) = (X^T X)^{-1} X\beta = \beta.$$

$$E(\hat{\beta}) = E(E(\hat{\beta}|X)) = E(\beta) = \beta.$$

$\Rightarrow$ unbiased estimator of β.

b) $Var(y|X) = Var(X\beta + u | X) = Var(u|X) = \sigma_u^2.$

$$Var(\hat{y}|X) = Var(X(X^T X)^{-1} X^T \beta / X) =$$

$$= X(X^T X)^{-1} X^T Var(\beta|X)(X(X^T X)^{-1} X^T)^T =$$

$$= \sigma^2 X(X^T X)^{-1} X^T w X (X^T X)^{-1} X^T.$$

d) $Cov(y, \hat{\beta}|X) = E(y|X) \cdot E(\hat{\beta}^T|X) - E(y \cdot \beta^T|X) =$

$$= E(X\beta + u|X) E(\hat{\beta}^T|X) - E((X\beta+u)\hat{\beta}^T|X) =$$

$$= E(X\beta|X) E(\hat{\beta}^T|X) + E(u|X) E(\beta^T|X) -$$

$$- E(X\beta\hat{\beta}^T|X) - E(u\hat{\beta}^T|X).$$

$$E(u\hat{\beta}^T|X) = -Cov(u, \hat{\beta}|X) + \overbrace{E(u|X)}^{0} E(\hat{\beta}|X).$$

$$E(u\hat{\beta}^T|X) = -Cov(u, \hat{\beta}|X).$$

$\mathrm{Cov}(y, \hat{\beta} | X) = \mathrm{Cov}(u, \hat{\beta} | X).$

1. We have two absolutely identical preliminary standardized regressors x and x. The dependent variable y is centered.

   In the ridge regression one minimizes the loss function

$$\text{loss}(\hat{\beta}) = (y - \hat{y})^T (y - \hat{y}) + \lambda \hat{\beta}^T \hat{\beta}, \quad \hat{y} = \hat{\beta}_1 x + \hat{\beta}_2 x.$$

   (a) Find the optimal $\hat{\beta}_1$ and $\hat{\beta}_2$ for fixed λ.

   (b) What happens to the estimates when $\lambda \to \infty$?

   (c) What happens to the sum $\hat{\beta}_1 + \hat{\beta}_2$ when $\lambda \to 0$?

a) $L(\hat{\beta}) = (y^T - \hat{y}^T)(y - \hat{y}) + \lambda (\hat{\beta})^T \hat{\beta} =$

$= y^T y - y^T X \hat{\beta} - \hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta} + \lambda \hat{\beta}^T \hat{\beta} \to \min_{\beta}$

FOC:

$d L(\hat{\beta}) = -y^T X \, d\hat{\beta} - d(\hat{\beta}^T) X^T y + d(\hat{\beta}^T (X^T X + \lambda I) \hat{\beta}) = 0$

$-2 y^T X \, d\hat{\beta} + 2 \hat{\beta}^T (X^T X + \lambda I) \, d\hat{\beta} = 0$

$\hat{\beta}^T (X^T X + \lambda I) - y^T X = 0$

$\hat{\beta} = (X^T X + \lambda I)^{-1} \cdot X y^T$

4. The columns of X are standardized. You know the SVD of the matrix $X = UDV^T$. The diagonal elements of D are positive and ordered from highest to lowest, $d_{11} > d_{22} > \cdots > 0$.

   Let's maximize $\|Xw\|^2$ by choosing an optimal vector w subject to $\|w\|^2 = 1$.

   (d) Write the Lagrangian function for this problem.

   (e) Find the first order conditions. Differential is your friend!

   (f) Find the optimal w in terms of columns of V.

   Hint: one may interpret the FOC in terms of eigenvalues and eigenvectors!

a) $X = U D V^T = U \begin{pmatrix} d_{11} & 0 & \cdots & \\ \vdots & d_{12} & \ddots & \\ 0 & & & 0 \end{pmatrix} V^T.$

$$\begin{cases} \|Xw\|^2 \to \max\limits_{w}. \\ \text{s.t. } \|w\|^2 = 1. \end{cases}$$

Let $w = \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix}$ $\quad X = \begin{pmatrix} x_{11} & \cdots & \\ \vdots & x_{22} & \\ \vdots & & \ddots & \\ & & & x_{nk} \end{pmatrix}$

$$\|w\| = \sqrt{w_1^2 + w_2^2 + \ldots + w_k^2}$$

$$\|w\|^2 = w_1^2 + w_2^2 + \ldots + w_k^2.$$

$$\|w\|^2 = w^T w = 1.$$

$$Xw = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & & \\ \vdots & & \ddots & \\ x_{n1} & & & x_{nk} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_k \end{pmatrix} =$$

$$= \begin{pmatrix} x_{11} w_1 + x_{12} w_2 + \ldots + x_{1k} w_k \\ \vdots \\ \vdots \end{pmatrix}$$

$$\|Xw\|^2 = \left(\sum x_{1i} w_i\right)^2 + \ldots + \left(\sum x_{ni} w_i\right)^2.$$

$$\|Xw\|^2 = (Xw)^T Xw.$$

$$\begin{cases} \|Xw\|^2 \to \max\limits_{w}. \\ s.t\ \|w\|^2 = 1. \end{cases}$$

$$\begin{cases} \sum \left(X_{1i} w_i\right)^2 + \dots + \left(\sum X_{ki} w_i^2\right) \to \max\limits_{w_i}. \\ s.t.\ \sum (w_i^2) = 1. \end{cases}$$

$$L = \|Xw\|^2 + \lambda\left(1 - \|w\|^2\right) = \left(\sum X_{1i} w_i\right)^2 + \dots + \left(\sum X_{ki} w_i\right)^2 +$$
$$+ \lambda\left(1 - \sum (w_i)^2\right).$$