CZ4041: Tutorial Week 11&12

Due on April 8, 2021 at 8:30am

Assoc Prof Pan, Sinno Jialin - CS4

Pang Yu Shao
U1721680D

08/04/2021

Problem 1

Suppose a dataset of four 3-dimensional instances is shown in Table 1. Estimate the sample mean and covariance matrix (unbiased).

Table 1: Data set for Question 1.

ID	X_1	X_2	X_3
P1	3	5	-1
P2	-1	8	3
P3	2	-4	-4
P4	0	-1	-6

Solution

Calculate sample mean (unbiased):

$$\begin{aligned}\hat{\mu} &= \frac{1}{N}\sum_{i=1}^{N} x_i \\ &= \frac{1}{4}\begin{bmatrix}(3-1+2) & (5+8-4-1) & (3-1-4-6)\end{bmatrix} \\ &= \begin{bmatrix}1 & 2 & -2\end{bmatrix}\end{aligned}$$

Therefore, centered data matrix:

$$\tilde{X} = \begin{bmatrix} 2 & 3 & 1 \\ -2 & 6 & 5 \\ 1 & -6 & -2 \\ -1 & -3 & -4 \end{bmatrix}$$

Calculate sample covariance (unbiased):

$$\begin{aligned}\tilde{\Sigma} &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\hat{\mu})(x_i-\hat{\mu})^T \\ &= \frac{1}{3}\tilde{X}^T\tilde{X} \\ &= \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 & -1 \\ 3 & 6 & -6 & -3 \\ 1 & 5 & -2 & -4 \end{bmatrix}\begin{bmatrix} 2 & 3 & 1 \\ -2 & 6 & 5 \\ 1 & -6 & -2 \\ -1 & -3 & -4 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 10 & -9 & -6 \\ -9 & 90 & 57 \\ -6 & 57 & 46 \end{bmatrix} \\ &= \begin{bmatrix} 3.33 & -3 & -2 \\ -3 & 30 & 19 \\ -2 & 19 & 15.33 \end{bmatrix}\end{aligned}$$

Problem 2

Suppose a dataset of 5 1-dimensional instances is shown in Table 2. Use histogram estimator with an origin of 0 and a width of 3, naive estimator with a width of 3, and 3-NN estimator to estimate the density function $\hat{p}(x)$ and compute the value of $\hat{p}(2.6)$ at 2.6, respectively.

Table 2: Data set for Question 2.

P1	P2	P3	P4	P5
1.2	2	10	-6	3.5

Solution
For **histogram estimator**:
Window: $0 \leq x_i < 3$

$$\begin{aligned}\hat{p}(2.6) &= \frac{2}{5 * 3} \\ &= 0.133\end{aligned}$$

For **naive estimator**:
Window: $1.1 \leq x_i < 4.1$

$$\begin{aligned}\hat{p}(2.6) &= \frac{3}{5 * 3} \\ &= 0.2\end{aligned}$$

For **K-NN estimator**:
Distance from x=2.6:

P2: 0.6
P5: 0.9
P1: 1.4 (3rd nearest neighbour)
P3: 7.4
P4: 8.6

$$\begin{aligned}\hat{p}(2.6) &= \frac{3}{5 * (2 * 1.4)} \\ &= 0.214\end{aligned}$$

Problem 3

A dataset of five 4-dimensional instances is given in Table 1. Suppose a SVD is performed on the data matrix X (5-by-4) via $X = VDU^T$, and the matrices U, D and V are shown in Tables 2-4, respectively. Use principal component analysis to project the 5 datapoints in Table 1 to 2-dimensional space.

Table 1: Data set for Question 1.

Data Points	X_1	X_2	X_3	X_4
P1	2	4	1	3
P2	1	2	3	5
P3	-2	-4	-4	-1
P4	0	-1	-2	-6
P5	-1	-1	2	-1

Table 2: The matrix **V** (5-by-5) obtained by SVD ($\mathbf{X} = \mathbf{VDU}^\top$)

-0.4577	0.2550	-0.5536	0.4680	0.4472
-0.5612	-0.2150	0.1896	-0.6348	0.4472
0.4497	-0.7183	-0.2749	0.0787	0.4472
0.5206	0.6063	-0.1153	-0.3849	0.4472
0.0486	0.0720	0.7542	0.4730	0.4472

Table 3: The matrix **D** (5-by-4) obtained by SVD ($\mathbf{X} = \mathbf{VDU}^\top$)

10.9040	0	0	0
0	4.8385	0	0
0	0	3.3973	0
0	0	0	0.3867
0	0	0	0

Table 4: The matrix **U** (4-by-4) obtained by SVD ($\mathbf{X} = \mathbf{VDU}^\top$)

-0.2224	0.3430	-0.3302	-0.8508
-0.4880	0.5756	-0.4046	0.5166
-0.4479	0.2924	0.8400	-0.0911
-0.7154	-0.6823	-0.1473	-0.0309

Problem 3 continued on next page...

Solution

First, we center the datapoints in X

$$X = \begin{bmatrix} 2 & 4 & 1 & 3 \\ 1 & 2 & 3 & 5 \\ -2 & -4 & -4 & -1 \\ 0 & -1 & -2 & -6 \\ -1 & -1 & 2 & -1 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore, datapoints are already centered.

2. Calculate the covariancematrix $\tilde{\Sigma}$,

$$\begin{aligned} \tilde{\Sigma} &= \frac{1}{5-1} X^T X \\ &= U \tilde{D} U^T, \tilde{D} = \frac{1}{4} D^T D \end{aligned}$$

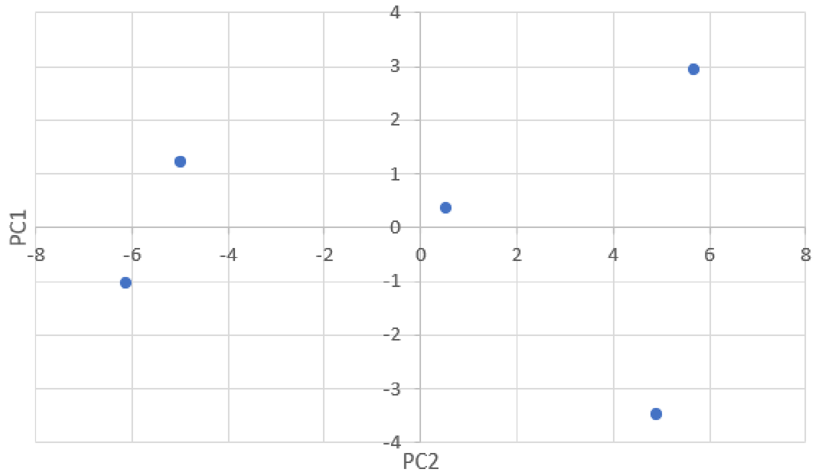
3. Find Eigenvalues/ Eigenvectors of the Covariance matrix.
Since $U^T U = I$,

$$\tilde{\Sigma} U = \tilde{D} U$$

$$\tilde{D} = \frac{1}{4} \begin{bmatrix} 10.90^2 & 0 & 0 & 0 \\ 0 & 4.84^2 & 0 & 0 \\ 0 & 0 & 3.40^2 & 0 \\ 0 & 0 & 0 & 0.39^2 \end{bmatrix}$$

Therefore, the first two columns of the U matrix correspond to the top two eigenvectors with the highest eigenvalues. We select the first 2 columns to construct the top 2 principal components.

$$\begin{aligned} XU &= \begin{bmatrix} 2 & 4 & 1 & 3 \\ 1 & 2 & 3 & 5 \\ -2 & -4 & -4 & -1 \\ 0 & -1 & -2 & -6 \\ -1 & -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} -0.2224 & 0.3430 \\ -0.4880 & 0.5756 \\ -0.4479 & 0.2924 \\ -0.7154 & -0.6823 \end{bmatrix} \\ &= \begin{bmatrix} -4.99 & 1.23 \\ -6.12 & -1.04 \\ 4.90 & -3.48 \\ 5.68 & 2.93 \\ 0.53 & 0.35 \end{bmatrix} \end{aligned}$$


4
3
2
1
0
-1
-2
-3
-4
-8
-6
-4
-2
0
2
4
6
8
PC1
PC2