6. Principal Component Analysis (PCA) - Theorem

Visualizing the Covariance Structure I

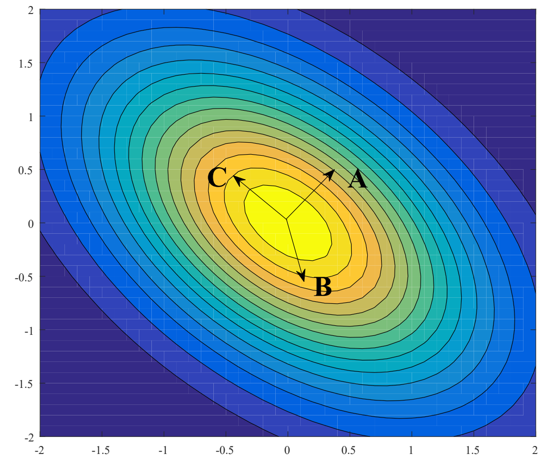
2/2 points (ungraded)

In this problem, we develop some intuition for PCA by visualizing the covariance structure of a random Gaussian vector

$$\mathbf{X} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}\right).$$

(In general, we will be considering very high-dimensional covariance structures, but here we will use a low-dimensional example for illustrative purposes.)

Consider the following image



Caption: In the above image, we graph the probability density function of $\mathbf{X}$ as follows. The point (x, y) is colored lighter if the density value at (x, y) is large and darker otherwise. The level curves consists of concentric ellipses centered at the origin $(0,0)^T$. The vector $\mathbf{A}$ points along the minor axis, the vector $\mathbf{C}$ points along the major axis, and the vector $\mathbf{B}$ points at a 135° angle (going clockwise) to $\mathbf{A}$. (Assume that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are all unit vectors.)

Along which direction is the variance of $\mathbf{X}$ the largest?

- A
- B
- C (selected)

✔

Along which direction is the variance of $\mathbf{X}$ the smallest?

A

B

C

Solution:

We start with the first question. The direction of maximal variance is that in which the ellipsoid is most stretched out, because this means that the distribution has more mass further away from the mean, which is the origin $(0,0)^T$. In other words, the direction in which the ellipses are most stretched out characterizes the directions which have the high probability of deviating far from the mean. Therefore, the direction we are looking for should be along the **major axis** of the ellipse. Hence, the correct answer is $\mathbf{C}$.

For the second question, we are looking for the direction of minimal variance. This will be the direction in which most of the mass is clustered near the mean, which is the origin $(0,0)^T$. Hence, the direction we are looking for is along the **minor axis** of the ellipse. The correct answer is thus $\mathbf{A}$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Visualizing the Covariance Structure II: Covariance Matrix of A Coordinate Transformation

4/4 points (ungraded)

Recall from the previous problem that $\mathbf{X} \sim N\left((0,0)^T, \Sigma\right)$ where

$$\Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}.$$

The covariance matrix of $\mathbf{X}$ can be decomposed as

$$\Sigma = PDP^T$$

where

$$P = \begin{pmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{pmatrix}, \quad D = \begin{pmatrix} 1.5 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

(It is not necessary for you to derive this.) In particular, the eigenvectors of Σ are the columns of P.

Consider the random vector $\mathbf{Y} = P^T\mathbf{X}$. It is true that

$$\mathbf{Y} = N\left((0,0)^T, \Sigma'\right)$$

for some matrix $\Sigma' \in \mathbb{R}^{2\times 2}$.

Fill in the entries of Σ' below.

Hint 1: The covariance matrix of $\mathbf{Y}$ is given by

$$\begin{aligned} \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^T\right] &= \mathbb{E}\left[\left(P^T\mathbf{X}\right)\left(P^T\mathbf{X}\right)^T\right] \\ &= P^T\mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]P. \end{aligned}$$

Hint 2: You need to use that $\Sigma = PDP^T$.

$(\Sigma')_{11} =$ 1.5 ✔ **Answer:** 1.5 $(\Sigma')_{12} =$ 0 ✔ **Answer:** 0

$(\Sigma')_{21} =$ 0 ✔ **Answer:** 0 $(\Sigma')_{22} =$ 0.5 ✔ **Answer:** 0.5

Solution:

As given in the hints, we know that the covariance matrix of $\mathbf{Y}$ is given by

$$\begin{aligned} \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^T\right] &= \mathbb{E}\left[\left(P^T\mathbf{X}\right)\left(P^T\mathbf{X}\right)^T\right] \\ &= P^T\mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]P \\ &= P^T\Sigma P \\ &= P^T\left(PDP^T\right)P \\ &= D. \end{aligned}$$

Therefore, $\Sigma' = D$ so that

$$\Sigma' = \begin{pmatrix} 1.5 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

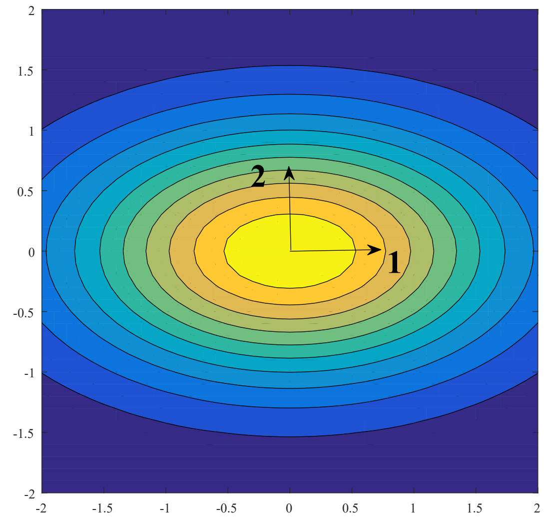
Visualizing the Covariance Structure III: Visualizing a Coordinate Transformation

2/2 points (ungraded)

As in the previous problem, $\mathbf{Y} = P^T\mathbf{X}$ where

$$\mathbf{X} \sim N\left((0,0)^T, \Sigma\right), \quad \Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}, \quad \Sigma = PDP^T.$$

Consider the graph below of the pdf of the Gaussian vector $\mathbf{Y}$.



Caption: The graph above represents the density of the random vector $\mathbf{Y}$. The point (x, y) is colored lighter if the density value at (x, y) is large and darker otherwise. The level curves consists of concentric ellipses centered at the origin $(0,0)^T$, with minor axis along the y-axis and major axis along the x-axis. The vector labeled $\mathbf{1}$ is given by $(1,0)^T$ and the vector labeled $\mathbf{2}$ is given by $(0,1)^T$.

Let $\mathbf{A}$ and $\mathbf{B}$ denote the vectors in the image for the problem "Visualizing the Covariance Structure I".

Which vector is equal to $P^T\mathbf{A}$?

- [ ] $\mathbf{1}$
- [x] $\mathbf{2}$

✔

Which vector is equal to $P^T\mathbf{B}$?

- [x] $\mathbf{1}$
- [ ] $\mathbf{2}$

Solution:

By the solution to the problem "Visualizing the Covariance Structure I", we know that $\mathbf{A}$ is the (column) eigenvector of Σ that corresponds to the eigenvalue 0.5 and $\mathbf{B}$ is the (column) eigenvector of Σ that corresponds to the eigenvalue 1.5. Hence, we have

$$P = \begin{pmatrix} \mathbf{B} & \mathbf{A} \end{pmatrix}.$$

Recall that $P^T P = I_2$. Hence,

$$P^T \mathbf{A} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad P^T \mathbf{B} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Therefore, for the first question, $P^T\mathbf{A}$ is the vector denoted by $\mathbf{2}$. For the second question, $P^T\mathbf{B}$ is the vector denoted by $\mathbf{1}$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Spectral Decomposition and Principal Component Analysis

(Caption will be displayed when you start playing the video.)

0:00 / 0:00 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

PCA and a Coordinate Transformation I

1/1 point (ungraded)

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ denote a data set (or point cloud) consisting of i.i.d. vectors. Let S denote the empirical covariance matrix of this data set. Since S is symmetric and positive semidefinite, we may apply the decomposition theorem to write

$$S = PDP^T$$

where $P \in \mathbb{R}^{d\times d}$ satisfies $P^T P = PP^T = I_d$ and

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{pmatrix}$$

is a diagonal matrix with entries given by the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ of S.

Consider the new (vector-valued) data set $P^T\mathbf{X}_1, \ldots, P^T\mathbf{X}_n$ obtained by applying the orthogonal matrix P^T to the vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Which of the following is the empirical covariance of the new data set $P^T\mathbf{X}_1, \ldots, P^T\mathbf{X}_n$? (To simplify calculation, you should assume that the sample mean of the new data set $P^T\mathbf{X}_1, \ldots, P^T\mathbf{X}_n$ is the zero vector: $\frac{1}{n}\sum_{i=1}^n P^T\mathbf{X}_i = \mathbf{0} \in \mathbb{R}^d$.)

Hint: Recall that the empirical covariance matrix S' of a set of vectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ is given by

$$S' = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{Y}_i \mathbf{Y}_i^T\right) - \overline{\mathbf{Y}}_n \overline{\mathbf{Y}}_n^T$$

where $\overline{\mathbf{Y}}_n$ denotes the empirical mean of $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$.

- (x) D
- ( ) $P^T DP$
- ( ) PDP^T

DP^T

Solution:

The correct answer is the first choice, D. We demonstrate this with the following calculation. Let S' denote the empirical covariance of the 'transformed' data set $P^T\mathbf{X}_1, \ldots, P^T\mathbf{X}_n$. By definition of the empirical covariance, linearity, and the fact that $\Sigma = PDP^T$, we have

$$\begin{aligned} S' &= \frac{1}{n}\sum_{i=1}^{n} P^T\mathbf{X}_i (P^T\mathbf{X}_i)^T \\ &= \frac{1}{n}\sum_{i=1}^{n} P^T\mathbf{X}_i\mathbf{X}_i P \\ &= P^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i\right)P \\ &= P^T S P \\ &= P^T\left(PDP^T\right)P \\ &= D. \end{aligned}$$

Remark: Applying the orthogonal matrix P^T can be thought of as changing the coordinate system of our data set. This is intuitive if we think of P^T as matrix that, when applied to a vector $\mathbf{v}$, rotates $\mathbf{v}$ without changing its length.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

PCA and a Coordinate Transformation II

1/1 point (ungraded)

We use the statistical set-up from the previous problem. In particular, recall that P is the orthogonal matrix such that the empirical covariance matrix of the original data set $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is given by

$$S = PDP^T.$$

Let $\mathbf{Y}_i = P^T\mathbf{X}_i$ for $1 \le i \le n$. The sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \in \mathbb{R}^d$ consists of i.i.d. vectors, and let's assume for simplicity that the empirical mean satisfies $\overline{Y}_n = \mathbf{0} \in \mathbb{R}^d$.

Consider the one-dimensional data set

$$\mathbf{Y}^1 = \mathbf{Y}_1^1, \mathbf{Y}_2^1, \ldots, \mathbf{Y}_n^1$$

consisting of the first coordinates of $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ and the data set

$$\mathbf{Y}^2 = \mathbf{Y}_1^2, \mathbf{Y}_2^2, \ldots, \mathbf{Y}_n^2$$

consisting of the second coordinates of $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$.

What is the empirical covariance between the data sets $\mathbf{Y}^1$ and $\mathbf{Y}^2$?

Hint: Recall that you computed the empirical covariance matrix of $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ in the previous question.

0

✔ **Answer:** 0

Solution:

As demonstrated in the previous problem, the empirical covariance matrix S' of the data set $\mathbf{Y}_1 = P^T\mathbf{X}_1, \ldots, P^T\mathbf{X}_n$ is given by

$$S' = D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{pmatrix}.$$

Recall that the i, j-th entry of S' gives the empirical covariance between the data sets

$$\mathbf{Y}^i = \mathbf{Y}^i_1, \mathbf{Y}^i_2, \ldots, \mathbf{Y}^i_n$$

and

$$\mathbf{Y}^j = \mathbf{Y}^j_1, \mathbf{Y}^j_2, \ldots, \mathbf{Y}^j_n.$$

Since $(S')_{ij} = 0$ if $i \neq j$ (*i.e.*, the matrix S' is a diagonal matrix, it follows that the empirical covariance between the real-valued data sets $\mathbf{Y}^1$ and $\mathbf{Y}^2$ is 0.

Remark: Applying the matrix P^T to our data set changes coordinates so that the transformed data set $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ has the property that the coordinate data sets $\mathbf{Y}^1, \ldots, \mathbf{Y}^d$ are **uncorrelated**. In this transformed data set, the principal axes are given by the span of the standard basis vectors $\{\mathbf{e}_i\}_{i=1}^d$, where $\mathbf{e}_i$ is the vector such that

$$\mathbf{e}_i^j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

In summary, applying P^T to the original data set $\mathbf{X}_1, \ldots, \mathbf{X}_n$ puts the data in a convenient standard form for carrying out PCA.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: (Optional) Unit 8 Principal component analysis:(Optional) Lecture 23: Principal Component Analysis / 6. Principal Component Analysis (PCA) - Theorem

Add a Post

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.

✖