DLI Accelerated Data Science Teaching Kit

Lecture 3.6 - Feature Reduction: PCA

Feature Reduction

Feature Reduction is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains meaningful properties of the original data.

- It is to reduce large numbers of observations and/or large numbers of variables for tasks such as signal processing, speech recognition, natural language processing, and bioinformatics.
- The difference between feature selection and feature reduction is that feature selection is directly to eliminate original features while feature reduction is to map the original high-dimensional feature space to a low-dimensional space.
- The data transformation may be linear or nonlinear.

Feature Reduction

A linear or non-linear transform on the original feature space

- Unsupervised methods
    - Principal Component Analysis (PCA)
    - Independent Component Analysis (ICA)
    - Autoencoder
- Supervised methods
    - Linear Discriminant Analysis (LDA)

Feature Reduction

A linear or non-linear transform on the original feature space.

- Unsupervised methods
  - Principal Component Analysis (PCA)
  - Independent Component Analysis (ICA)
  - Autoencoder
- Supervised methods
  - Linear Discriminant Analysis (LDA)

PCA (Principal component analysis)

PCA is to calculate the principal components for data analytics, where only the a few of principal components are selected in terms of a predefined number K of features.

- It is commonly used for dimensionality reduction by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible.

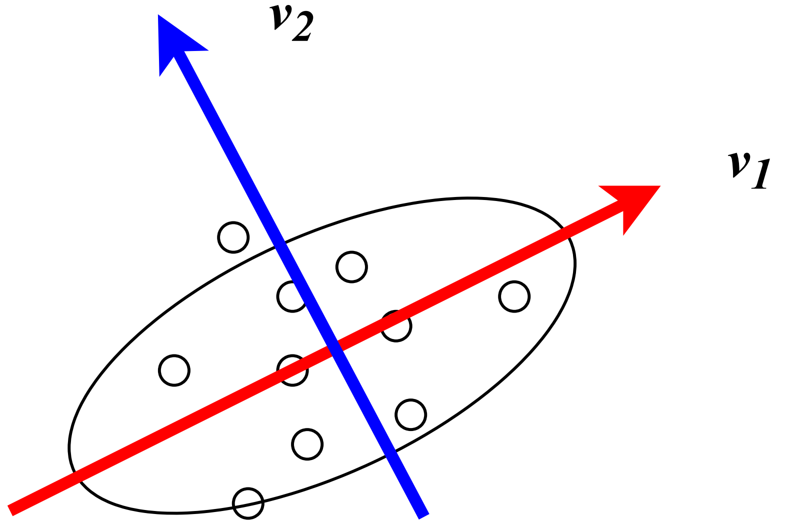
- The first principal component can equivalently be defined as a direction that maximizes the variance of the projected data.

- The $\boldsymbol{i}^{\mathbf{th}}$ principal component can be taken as a direction orthogonal to the $\boldsymbol{i-1}$ principal components, which maximizes the variance of the projected data.

PCA (Principal component analysis)

The goal of PCA is to reduce the dimensionality of the data while preserving the variation present in the dataset as much as possible.

- The axes have been rotated to new principal components such that:
  - Principal component 1 has the highest variance.
  - Principal component 2 has the next highest variance, and so on.



v_1: **Principal Component 1**

v_2: **Principal Component 2**

PCA (Principal component analysis)

Principal Components (PCs): orthogonal vectors that are ordered by the fraction of the total information (variation) in the corresponding directions.

PCs can be found as the “best” eigenvectors of the covariance matrix of the data points.

PCs are linear least squares fits to samples, each orthogonal to the previous PCs:

- First PC is a minimum distance fit to a vector in the original feature space
- Second PC is a minimum distance fit to a vector in the plane perpendicular to the first PC

Mathematics for PCA

Covariance Matrix

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix} = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_d) \end{bmatrix}$$

$$\boldsymbol{\Sigma} = E[(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})^T]$$

Estimating covariance matrix from data points $\{x^{(i)}\}_{i=1}^{N}$:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{N}(\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}})$$

$$\widetilde{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}^{(1)} - \widehat{\boldsymbol{\mu}} \\ \vdots \\ \boldsymbol{x}^{(N)} - \widehat{\boldsymbol{\mu}} \end{bmatrix} \qquad \widehat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}^{(i)}$$

Mean-centered data

Calculating PCA

Input: $\boldsymbol{N}\times\boldsymbol{d}$ data matrix X (each row contains a d dimensional data points)

$$\widetilde{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}^{(1)} - \widehat{\boldsymbol{\mu}} \\ \vdots \\ \boldsymbol{x}^{(N)} - \widehat{\boldsymbol{\mu}} \end{bmatrix}$$

Mean-centered data

- $\boldsymbol{\mu} = \frac{\boldsymbol{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{N}} \boldsymbol{x}^{(\boldsymbol{i})}$ **// calculate the mean of data**
- $\tilde{X} \leftarrow$ Mean values of data points is subtracted from rows of data.
- $\sum = \frac{1}{N} \tilde{X}^T \tilde{X}$ (Covariance matrix) // **calculate the covariance of data**
- Calculate eigenvalues and eigenvectors of the covariance matrix
- Pick eigenvectors corresponding to the largest eigenvalues based on a predefined dimensions K and put them in the columns of $\boldsymbol{e} = [\boldsymbol{e_1}, \ \dots, \boldsymbol{e_{d'}}]$
    - $\boldsymbol{K} < \boldsymbol{d}$
- $\boldsymbol{X'} = \boldsymbol{eT}\tilde{X}$ (Obtaining PCs)

Output: $\boldsymbol{X'}$ (PCs)

An Example of PCA application

House Information

#	House Price	Area
1	10	9
2	2	3
3	1	2
4	7	6.5
5	3	2.5

- Two features of house data
    - House Price (Million of dollars)
    - House Area

An Example of PCA application

Calculate $\boldsymbol{\mu} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{x}^{(\boldsymbol{i})}$

#	House Price	Area
1	10	9
2	2	3
3	1	2
4	7	6.5
5	3	2.5

$$\boldsymbol{\mu}_{\text{House Price}} = \frac{10 + 2 + 1 + 7 + 3}{5} = 4.6$$

$$\boldsymbol{\mu}_{\text{Area}} = \frac{9 + 3 + 2 + 6.5 + 2.5}{5} = 4.6$$

An Example of PCA application

Calculate $\widetilde{X} \leftarrow$ Mean values of data points is subtracted from rows of data.

#	House Price	Area
1	10 – 4.6	9 – 4.6
2	2 – 4.6	3 – 4.6
3	1 – 4.6	2 – 4.6
4	7 – 4.6	6.5 – 4.6
5	3 – 4.6	2.5 – 4.6

#	House Price	Area
1	5.4	4.4
2	-2.6	-1.6
3	-3.6	-2.6
4	2.4	1.9
5	-1.6	-2.1

An Example of PCA application

Calculate $\sum = \frac{1}{N}\tilde{X}^T\tilde{X}$ (Covariance matrix)

#	House Price	Area
1	5.4	4.4
2	-2.6	-1.6
3	-3.6	-2.6
4	2.4	1.9
5	-1.6	-2.1

$$A = \begin{pmatrix} 5.4 \\ -2.6 \\ -3.6 \\ 2.4 \\ -1.6 \end{pmatrix} \qquad B = \begin{pmatrix} 4.4 \\ -1.6 \\ -2.6 \\ 1.9 \\ -2.1 \end{pmatrix}$$

$$\sum = \begin{pmatrix} Var(A) & Cov(A,B) \\ Cov(A,B) & Var(B) \end{pmatrix} = \begin{matrix} Cov(A,A) & Cov(A,B) \\ Cov(A,B) & Cov(B,B) \end{matrix} = \frac{1}{5}\begin{pmatrix} A \cdot A & A \cdot B \\ A \cdot B & B \cdot B \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 57.2 & 45.2 \\ 45.2 & 36.7 \end{pmatrix}$$

An Example of PCA application

Calculate eigenvalues $\boldsymbol{\lambda}$ and eigenvectors $\boldsymbol{e}$ of the covariance matrix with **singular value decomposition (SVD).**

#	House Price	Area
1	5.4	4.4
2	-2.6	-1.6
3	-3.6	-2.6
4	2.4	1.9
5	-1.6	-2.1

$$\Sigma = \begin{pmatrix} -0.78 & -0.62 \\ -0.62 & 0.78 \end{pmatrix} \begin{pmatrix} 18.66 & 0 \\ 0 & 0.12 \end{pmatrix} \begin{pmatrix} -0.78 & -0.62 \\ -0.62 & 0.78 \end{pmatrix}$$

$$e_1 = \begin{pmatrix} -0.78 \\ -0.62 \end{pmatrix}. \quad e_2 = \begin{pmatrix} -0.62 \\ 0.78 \end{pmatrix}. \quad \lambda = \begin{pmatrix} 18.66 \\ 0.12 \end{pmatrix}$$

An Example of PCA application

Pick eigenvectors corresponding to the largest eigenvalues and put them in the columns of $\boldsymbol{e} = [\boldsymbol{e_1}, \dots, \boldsymbol{e_{d'}}]$

#	House Price	Area
1	5.4	4.4
2	-2.6	-1.6
3	-3.6	-2.6
4	2.4	1.9
5	-1.6	-2.1

$$e_1 = \begin{pmatrix} -0.78 \\ -0.62 \end{pmatrix}. \quad e_2 = \begin{pmatrix} -0.62 \\ 0.78 \end{pmatrix}.$$

$$e = (e_1, e_2) = \begin{pmatrix} -0.78 & -0.62 \\ -0.62 & 0.78 \end{pmatrix}.$$

An Example of PCA application

Obtain PCs $\boldsymbol{X'} = \boldsymbol{eT}\tilde{X}$

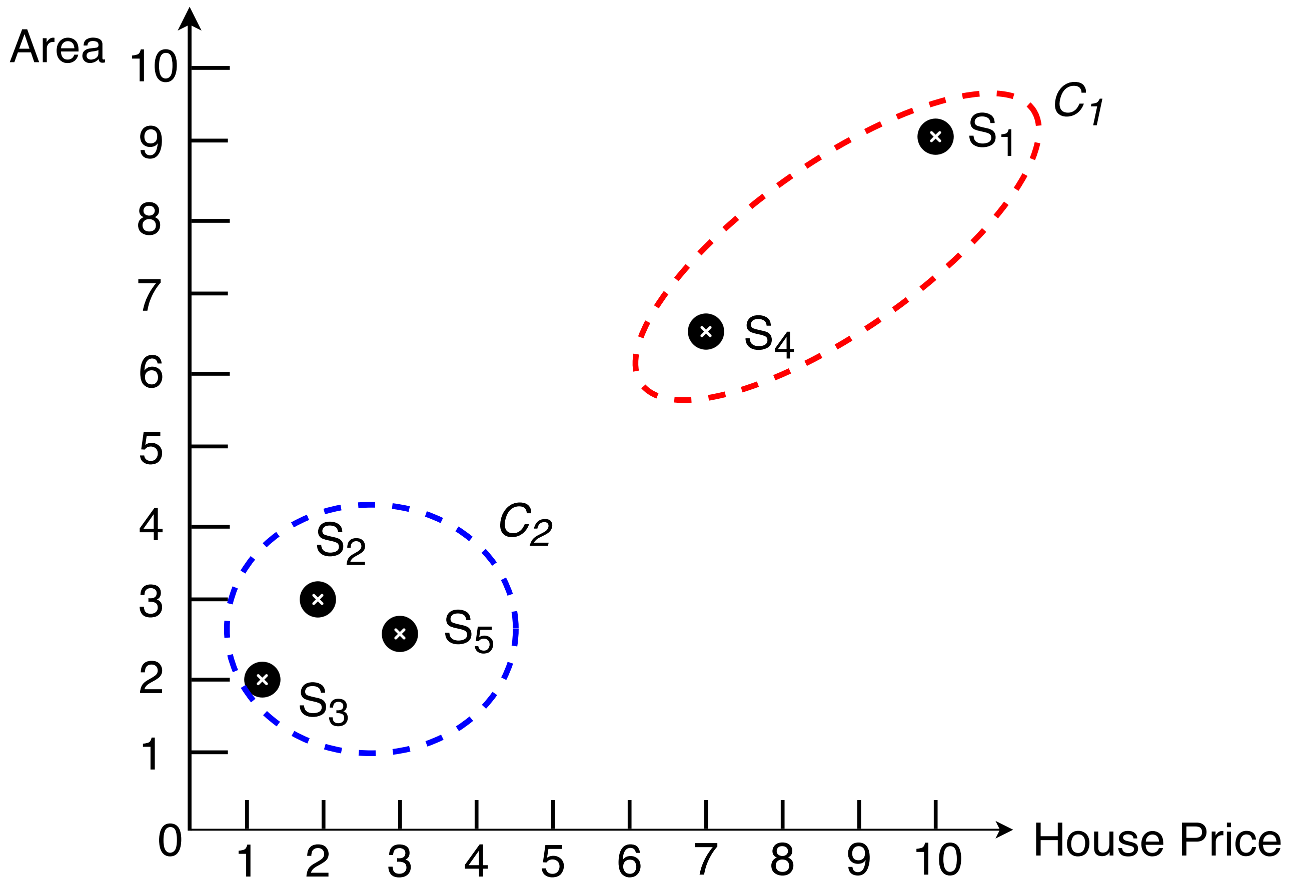
$$\tilde{X} = \begin{pmatrix} 5.4 & 4.4 \\ -2.6 & -1.6 \\ -3.6 & -2.6 \\ 2.4 & 1.9 \\ -1.6 & -2.1 \end{pmatrix}$$

$$e = (e_1, e_2) = \begin{pmatrix} -0.78 & -0.62 \\ -0.62 & 0.78 \end{pmatrix}.$$

$$-6.94 = -0.78 \times 5.4 + 4.4 \times (-0.62)$$
$$0.084 = -0.62 \times 5.4 + 0.78 \times 4.4$$

#	PC1	PC2
1	-6.94	0.084
2	3.02	0.364
3	4.42	0.204
4	-3.05	-0.006
5	2.55	-0.646

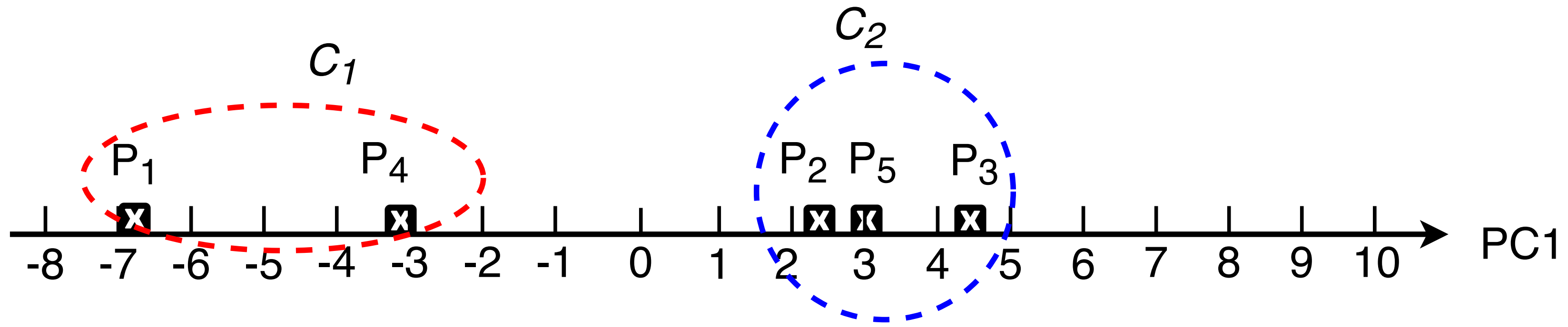
PCA for clustering



Point	House Price	Area
S_1	10	9
S_2	2	3
S_3	1	2
S_4	7	6.5
S_5	3	2.5

PCA

Map 2D to 1D



Point	PC1
P_1	-6.94
P_2	3.02
P_3	4.42
P_4	-3.05
P_5	2.55

DLI Accelerated Data Science Teaching Kit

Thank You