Exercise 5 Dimensionality Reduction Introduction

Matthis Brocheton

Probleme 1

The general form of data mapping from high to low-dimensionality space according to PCA:

$$Z = (X - \mu)W$$

$\boldsymbol{X} \in \mathbb{R}^{\mathbf{nxd}}$: the initial data matrix with **n** samples and **d** variables.

$\boldsymbol{\mu} \in \mathbb{R}^{\mathbf{d}}$: the mean vector of the columns of **X**. This vector is subtracted to center the data.

$\boldsymbol{X} - \boldsymbol{\mu}$: the centered data, meaning the data without the bias introduced by the mean.

$\boldsymbol{W} \in \mathbb{R}^{\mathbf{dxk}}$: the projection matrix, composed of the **k** eigenvectors (called principal components) associated with the largest eigenvalues of the covariance matrix.

$\boldsymbol{Z} \in \mathbb{R}^{\mathbf{nxk}}$: the new projected data, now in a reduced **k**-dimensional space (with **k < d**).

Orthogonality

$\mathrm{W^T W = I_k}$

Each vector points in a unique direction.

No principal component "absorbs" or overlaps the information of another.

Matrix:

$$X = \begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 3 & 5 \\ 5 & 4 \\ 5 & 6 \\ 6 & 5 \\ 8 & 7 \\ 9 & 8 \end{bmatrix}$$

Mean data :

$\hat{x} = 5$

$\hat{y} = 5$

Center data:

$X_{center} = X - 1_n \hat{x}^T$

$$X_{center} = \begin{bmatrix} -4 & -3 \\ -2 & -2 \\ -2 & 0 \\ 0 & -1 \\ 0 & 1 \\ 1 & 0 \\ 3 & 2 \\ 4 & 3 \end{bmatrix}$$

Calculation of the covariance matrix:

$$C = \frac{1}{n-1} X^T X$$

$$C = \frac{1}{7} \begin{bmatrix} -4 & -2 & -2 & 0 & 0 & 1 & 3 & 4 \\ -3 & -2 & 0 & -1 & 1 & 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} -4 & -3 \\ -2 & -2 \\ -2 & 0 \\ 0 & -1 \\ 0 & 1 \\ 1 & 0 \\ 3 & 2 \\ 4 & 3 \end{bmatrix}$$

$$\mathrm{Var(x)} = \frac{7}{1}((-4)^2 + (-2)^2 + (-2)^2 + 0^2 + 0^2 + 1^2 + 3^2 + 4^2$$
$$= 71(16 + 4 + 4 + 0 + 0 + 1 + 9 + 16) = \frac{50}{7}$$

$$Var(y) = 71((-3)2 + (-2)2 + 02 + (-1)2 + 12 + 02 + 22 + 32)$$
$$= 71(9 + 4 + 0 + 1 + 1 + 0 + 4 + 9) = \frac{28}{7} \ = 4$$

$$Cov(x, y) = \frac{7}{1}((-4)(-3) + (-2)(-2) + (-2)(0) + (0)(-1) + (0)(1) + (1)(0) + (3)(2)$$
$$+ (4)(3)) = 71(12 + 4 + 0 + 0 + 0 + 0 + 6 + 12) = \frac{34}{7}$$

Covariance matrix :

$$C = \begin{bmatrix} \frac{50}{7} & \frac{34}{7} \\ \frac{34}{7} & 4 \end{bmatrix} \approx \begin{bmatrix} 7.14 & 4.86 \\ 4.86 & 4 \end{bmatrix}$$

Now I want to find Eigenvalues:

$$det(C - \lambda I) = \begin{bmatrix} 7.14 - \lambda & 4.86 \\ 4.86 & 4.00 - \lambda \end{bmatrix}$$

$$\det = (7.14 - \lambda)(4.00 - \lambda) - (4.86)^2$$

$$(7.14 - \lambda)(4.00 - \lambda) = 7.14 \cdot 4.00 - (7.14 + 4.00)\lambda + \lambda 2 = 28.56 - 11.14\lambda + \lambda 2$$

$$\lambda 2 - 11.14\lambda + 28.56 - 23.6196 = 0$$

$$\lambda 2 - 11.14\lambda + 4.9404 = 0$$

Quadratic formula

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$a = 1$$

$$b = -11.14$$

$$c = 4.9404$$

$$\lambda = \frac{11.14 \pm \sqrt{-11.14^2 - 4 * 1 * 4.9404}}{2 * 1} \approx 10.21$$

result for Eigenvalues λ

$$\lambda_1 = \frac{11.14 + 10.21}{2} = \frac{21.35}{2} = 10.675$$

$$\lambda_2 = \frac{11.14 - 10.21}{2} = \frac{0.93}{2} = 0.465$$

Now I want to find the first eigenvector:

For $\lambda_1 = 10.675$

$$C - \lambda_1 I = \begin{bmatrix} 7.14 - 10.675 & 4.86 \\ 4.86 & 4.00 - 10.675 \end{bmatrix} = \begin{bmatrix} -3.535 & 4.86 \\ 4.86 & -6.675 \end{bmatrix}$$

We are looking for vectors:

$$\begin{bmatrix} x \\ y \end{bmatrix}$$

$$-3.535x + 4.86y = 0$$

Calculation

$4.86y = 3.535x$

$y = \frac{3.535}{4.36}x$

$y = 0.7275x$

$v_1 = \begin{bmatrix} x \\ 0.7275x \end{bmatrix}$

If $x = 1$ then:

$v_1 = \begin{bmatrix} 1 \\ 0.7275 \end{bmatrix}$

Normalization:

$\| v1 \| = \sqrt{1^2 + (0.7275)^2} \approx 1.236$

$v_1 = \frac{1}{1.236} \begin{bmatrix} 1 \\ 0.7275 \end{bmatrix} \approx \begin{bmatrix} 0.809 \\ 0.589 \end{bmatrix}$

Result for the first eigenvector

$v_1 = \begin{bmatrix} 0.809 \\ 0.589 \end{bmatrix}$

Now I want to find the second eigenvector:

For $\lambda_2 = 0.465$

$C - \lambda_1 I = \begin{bmatrix} 7.14 - 0.465 & 4.86 \\ 4.86 & 4.00 - 0.465 \end{bmatrix} = \begin{bmatrix} 6.675 & 4.86 \\ 4.86 & 3.535 \end{bmatrix}$

We are looking for vectors:

$\begin{bmatrix} x \\ y \end{bmatrix}$

$6.675x + 4.86y = 0$

Calculation

$y = -\frac{6.675}{4.86}x$

$y = -1.374x$

$v_2 = [\genfrac{}{}{0pt}{}{x}{-1.374x}]$

If $x = 1$ then:

$v_2 = [\genfrac{}{}{0pt}{}{1}{-1.374}]$

Normalization:

$\| v2 \| = \sqrt{1^2 + (-1.374)^2} \approx 1.7$

$v_2 = \frac{1}{1.7} [\genfrac{}{}{0pt}{}{1}{-1.374}] \approx [\genfrac{}{}{0pt}{}{0.589}{0.809}]$

Result second *eigenvector*:

$v_2 = \begin{bmatrix} -0.589 \\ 0.809 \end{bmatrix}$

Result we can see that the v_1 *and* v_2 are orthogonal:

$v_1 = [\genfrac{}{}{0pt}{}{0.809}{0.589}]$

$v_2 = \begin{bmatrix} -0.589 \\ 0.809 \end{bmatrix}$

Probleme 2

a)

Between-class scatter

$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$

Within-class scatter

$S_W = \Sigma_1 + \Sigma_2$

b)

$$J(\omega) = \frac{(\omega^T(\mu_1 - \mu_2))^2}{\omega^T(\Sigma_1 + \Sigma_2)\omega}$$

maximized for:

$\omega = ((\Sigma1+\Sigma2)^{-1}(\mu1-\mu2))$

<u>The goal:</u>

Maximize the distance between the projected averages

Minimize the total variance after projection

Formula projection:

$y = \omega^T x$

Then:

The projected mean is $\omega_i: \mu_{yi} = \omega^T \mu_i$

The projected variance is $\sigma_i^2 = \omega^T \Sigma_i \omega$

So the Fisher criterion becomes:

$$J(\omega) = \frac{\left(\omega^T(\mu1 - \mu2)\right)^2}{\omega^T(\Sigma1 + \Sigma2)\omega}$$

This is a Rayleigh quotient

$$J(\omega) = \frac{\omega^T A \omega}{\omega^t B \omega}$$

$A = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ -> Between-class scatter

$B = \Sigma_1 + \Sigma_2$ -> Within-class scatter

Maximize this quotient

We want:

$$arg\omega max J(\omega) = \frac{\left(\omega^T(\mu_1 - \mu_2)\right)^2}{\omega^T(\Sigma 1 + \Sigma 2)\omega}$$

We have:

$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$

$S_W = \Sigma_1 + \Sigma_2$

So:

$$J(\omega) = \frac{\omega^T S_B \omega}{\omega^T S_W \omega}$$

Derive and cancel the gradient

We are maximizing a quadratic ratio function → we use Lagrange method.

the maximum of:

$$J(\omega) = \frac{\omega^T S_B \omega}{\omega^T S_W \omega}$$

Is:

$$\omega = S_W^{-1}(\mu_1 - \mu_2)$$

Because the derivative of $J(\omega)$ with respect to ω gives:

$$\frac{\partial J(\omega)}{\partial \omega} = 0 \rightarrow \ \omega \propto S_W^{-1}(\mu_1 - \mu_2)$$

Conclusion:

We have shown that the optimal projection vector ω is:

$\omega = (\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2)$

It is the direction that maximizes class separation while minimizing projected total variance.