

DD2434 - Machine Learning, Advanced Course
Assignment 2A

Tristan Perrot
tristanp@kth.se

Étienne Riguet
riguet@kth.se

December 2023



Contents

1	Principal Component Analysis	3
1.1	Question 1	3
1.2	Question 2	3
1.3	Question 3	3
1.4	Question 4	3
1.5	Question 5	4
2	PCA vs. Johnson-Lindenstrauss random projections	4
3	Programming task — MDS	4
A	Appendix	5

1 Principal Component Analysis

1.1 Question 1

Centering the data is a crucial step in Principal Component Analysis (PCA) because it removes the mean bias and makes the interpretation of principal components more straightforward. Indeed, if the data is not centered, the first principal component will be the direction of the mean of the data, and the second principal component will be the direction of the variance of the data. Therefore, the first principal component will not be the direction of the maximum variance of the data, and the second principal component will not be the direction of the second maximum variance of the data. In other words, the principal components will not be the directions of the maximum variance of the data.

1.2 Question 2

TODO

1.3 Question 3

The *variance* of the dataset \mathcal{Y} is defined as $\text{Var}(\mathcal{Y}) = \sum_{y \in \mathcal{Y}} \|y - \bar{y}\|_2^2 = \sum_{y \in \mathcal{Y}} \|y\|_2^2$ here because the data is centered. Therefore we have:

$$\begin{aligned}
 \text{Var}(\mathcal{Y}) &= \sum_{y \in \mathcal{Y}} \|y\|_2^2 = \sum_{y \in \mathcal{Y}} y^T y \\
 &= \sum_{i=1}^d (Y^T Y)_{i,i} \\
 &= \text{Tr}(Y^T Y) \\
 &= \text{Tr}(V \Sigma^T U^T U \Sigma V^T) \\
 &= \text{Tr}(V \Sigma^T \Sigma V^T) \\
 &= \text{Tr}(\Sigma^T \Sigma) \\
 \text{Var}(\mathcal{Y}) &= \sum_{i=1}^d \sigma_i^2
 \end{aligned} \tag{1}$$

1.4 Question 4

The *variance* of the projected dataset after PCA is $\text{Var}(\mathcal{X}) = \sum_{x \in \mathcal{X}} \|x\|_2^2$. Where $X = W^T Y$ is a $k \times n$ projected matrix. Therefore we have:

$$\begin{aligned}
 \text{Var}(\mathcal{X}) &= \sum_{x \in \mathcal{X}} \|x\|_2^2 = \sum_{x \in \mathcal{X}} x^T x \\
 &= \sum_{i=1}^k (X^T X)_{i,i} \\
 &= \text{Tr}(X^T X) \\
 &= \text{Tr}(Y^T W W^T Y) \\
 &= \text{Tr}(W W^T Y Y^T) \\
 &= \text{Tr}(W W^T V \Sigma^T \Sigma V^T) \\
 &= \text{Tr}(W W^T \Sigma^T \Sigma) \\
 \text{Var}(\mathcal{X}) &= \sum_{i=1}^k \sigma_i^2
 \end{aligned} \tag{2}$$

1.5 Question 5

The residual data points are $\mathcal{Z} = z_1, \dots, z_n$ where $z_i = y_i - WW^T y_i$. Therefore we have:

$$\begin{aligned}
 \text{Var}(\mathcal{Z}) &= \sum_{z \in \mathcal{Z}} \|z\|_2^2 = \sum_{z \in \mathcal{Z}} z^T z \\
 &= \sum_{y \in \mathcal{Y}} (y - WW^T y)^T (y - WW^T y) \\
 &= \sum_{y \in \mathcal{Y}} y^T y - y^T WW^T y - (WW^T y)^T y + (WW^T y)^T WW^T y \\
 &= \sum_{y \in \mathcal{Y}} y^T y - y^T WW^T y - y^T WW^T y + y^T WW^T WW^T y \\
 &= \sum_{y \in \mathcal{Y}} y^T y - 2y^T WW^T y + y^T WW^T y \\
 &= \sum_{y \in \mathcal{Y}} y^T y - y^T WW^T y \\
 &= \sum_{i=1}^d \sigma_i^2 - \sum_{i=1}^k \sigma_i^2 \\
 \text{Var}(\mathcal{Z}) &= \sum_{i=k+1}^d \sigma_i^2
 \end{aligned} \tag{3}$$

And therefore we conclude that variance of original data = variance explained by PCA + variance of residual data.

2 PCA vs. Johnson-Lindenstrauss random projections

TODO

3 Programming task — MDS

TODO

A Appendix