DD2434 - Machine Learning, Advanced Course
Assignment 2A

Tristan Perrot
tristanp@kth.se

Étienne Riguet
riguet@kth.se

December 2023

Contents

Tristan Perrot
Étienne Riguet

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. PCA doesn't manage intercept so the principal components inevitably come through the origin and therefore the first principal component will be the "line" which passes through the mean and the origin.

1.2 Question 2

The previous argument implies that a single SVD operation is sufficient to perform PCA both on rows and the columns of a data matrix.

1.3 Question 3

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

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

1.4 Question 4

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

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

Tristan Perrot
Étienne Riguet

1.5 Question 5

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

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

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

Tristan Perrot
Étienne Riguet

A Appendix