Exercises
Computational Intelligence Lab
SS 2017

Machine Learning Institute
Dept. of Computer Science, ETH Zürich
Prof. Dr. Thomas Hofmann
Web http://cil.inf.ethz.ch/

Series 2, March 9-10, 2017 (Principal Component Analysis)

Problem 1 (PCA Theory):

1. (a) $\bar{\mathbf{X}} = \mathbf{X} - \mathbf{M}$

   (b) $\mathbf{\Sigma} = \frac{1}{N}\bar{\mathbf{X}}\bar{\mathbf{X}}^\top \in \mathbb{R}^{D\times D}$

   (c) $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$. In the sequel we assume that $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_D)$, where $\lambda_1 \geq \ldots \geq \lambda_D \geq 0$. The eigenvalues are positive because $\mathbf{\Sigma}$ is symmetric. Further, the eigenvector matrix $\mathbf{U}$ can be written as $\mathbf{U} = [u_1, \ldots, u_D]$, where $u_i \in \mathbb{R}^D$ are unit eigenvectors (i.e. $\|u_i\|_2 = 1$) represented as column vectors.

   (d) $\bar{\mathbf{Z}}_K = \mathbf{U}_K^\top \bar{\mathbf{X}}$. Here, we have $\mathbf{U}_K$ is given by the first K columns of $\mathbf{U}$, i.e. $\mathbf{U}_K = [u_1, \ldots, u_K]$.

   (e) $\tilde{\mathbf{X}} = \mathbf{U}_K \bar{\mathbf{Z}}_K$

   (f) We have that $\tilde{\mathbf{X}} = \mathbf{U}_K\mathbf{U}_K^\top\bar{\mathbf{X}}$. The reconstruction error is :

   $$\text{err} = \frac{1}{N}\sum_{i=1}^{N}\|\tilde{x_i} - \bar{x_i}\|_2^2 = \|\tilde{\mathbf{X}} - \bar{\mathbf{X}}\|_F^2 = \|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2$$

   where $\|A\|_F = \sqrt{\text{trace}(AA^\top)} = \sqrt{\sum_i \sigma_i^2}$ is the Frobenius norm of matrix A and σ_i are its singular values (the same as eigenvalues if A is symmetric). Thus,

   $$\begin{aligned}
   \text{err} &= \frac{1}{N}\text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\bar{\mathbf{X}}^\top(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
   &= \text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\mathbf{\Sigma}(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)) \\
   &= \text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)) \\
   &= \text{trace}((\mathbf{U}_K\mathbf{U}_K^\top\mathbf{U} - \mathbf{U})\mathbf{\Lambda}(\mathbf{U}^\top\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{U}^\top)) \\
   &= \text{trace}(([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})\mathbf{\Lambda}([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})^\top) \\
   &= \text{trace}(\sum_{i=K+1}^{D}\lambda_i u_i u_i^\top) \\
   &= \sum_{i=K+1}^{D}\lambda_i \cdot \text{trace}(u_i u_i^\top) \\
   &= \sum_{i=K+1}^{D}\lambda_i
   \end{aligned}$$

   where we used the fact that $\text{trace}(u_i u_i^\top) = \|u_i\|_2^2 = 1$.

2. (a) Intrinsic dimensionality: high
   No knee in eigenvalue spectrum

   (b) No, the approximation error is the sum of the discarded eigenvalues and λ_{100} is still large.

   (c) $D = 100$ (no reduction)

3.

1. $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ Answer: (B)

2. $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ Answer: (E)

3. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ Answer: (C)

4. (a) We would like to decouple the dimensions/measurements in the transformed dataset, i.e. we would like to have uncorrelated dimensions.

   (b) Consider $\mathbf{Z} = \mathbf{A}^\top \mathbf{X}$. Let $\bar{\mathbf{x}}$ be the mean of the dataset $\mathbf{X}$. We write $\mathbf{M_X} = \overbrace{[\bar{\mathbf{x}}, ..., \bar{\mathbf{x}}]}^{N \text{ times}}$, correspondingly, $\mathbf{M_Z} = A^\top \mathbf{M_X}$. We can write the covariance matrix of $\mathbf{X}$ as $\mathbf{\Sigma_X} = (\mathbf{X} - \mathbf{M_X})(\mathbf{X} - \mathbf{M_X})^\top$.
   The covariance of $\mathbf{Z}$ is then given by:

$$
\begin{aligned}
\mathbf{\Sigma_Z} &= (\mathbf{Z} - \mathbf{M_Z})(\mathbf{Z} - \mathbf{M_Z})^\top \\
&= (\mathbf{A}^\top \mathbf{X} - \mathbf{M_Z})(\mathbf{A}^\top \mathbf{X} - \mathbf{M_Z})^\top \\
&= (\mathbf{A}^\top \mathbf{X} - \mathbf{A}^\top \mathbf{M_X})(\mathbf{A}^\top \mathbf{X} - \mathbf{A}^\top \mathbf{M_X})^\top \\
&= \mathbf{A}^\top (\mathbf{X} - \mathbf{M_X})(\mathbf{A}^\top (\mathbf{X} - \mathbf{M_X}))^\top \\
&= \mathbf{A}^\top (\mathbf{X} - \mathbf{M_X})(\mathbf{X} - \mathbf{M_X})^\top \mathbf{A} \\
&= \mathbf{A}^\top \mathbf{\Sigma_X} \mathbf{A}
\end{aligned}
$$

   (c) If we use $\mathbf{A} = \mathbf{U}$, we obtain:

$$
\begin{aligned}
\mathbf{\Sigma_Z} &= \mathbf{A}^\top \mathbf{\Sigma_X} \mathbf{A} \\
&= \mathbf{U}^\top \mathbf{\Sigma_X} \mathbf{U} \\
&= \mathbf{U}^\top \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top \mathbf{U} \\
&= \mathbf{U}^{-1} \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{-1} \mathbf{U} \\
&= \mathbf{I} \mathbf{\Lambda} \mathbf{I} \\
&= \mathbf{\Lambda}
\end{aligned}
$$

   We see that the covariance matrix of $\mathbf{Z}$ becomes the diagonal eigenvalue matrix $\mathbf{\Lambda}$: Choosing the eigenvectors associated with the highest eigenvalues results in capturing high variances in the transformed dataset.