Dimensionality Reduction

Octavian Ganea, Yannic Kilcher

March 9-10, 2017

PCA Exercise 1

Question 1: Review the steps to perform PCA mathematically.

Focus: PCA for compression.

PCA Step-by-step I

Organize the Dataset: Write the data as a matrix $\mathbf{X}$ of $D \times N$: N instances of D dimensional data.

Calculate the Empirical Mean:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n$$

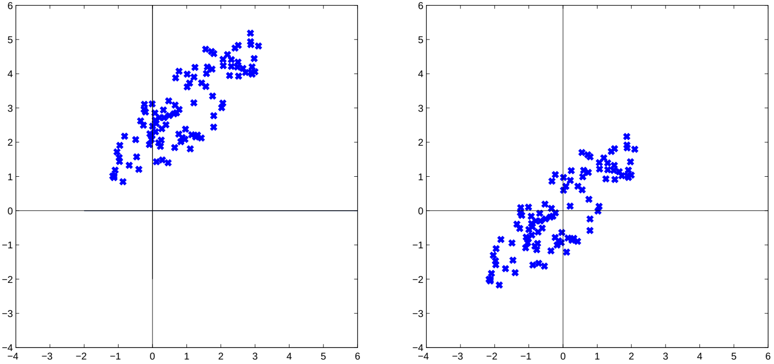
Center the data: Center the data by subtracting the mean from each data sample:

$$\bar{\mathbf{X}} = \mathbf{X} - \mathbf{M}$$

where $\mathbf{M} = \underbrace{[\bar{\mathbf{x}}, ..., \bar{\mathbf{x}}]}_{N \text{ times}}$

PCA Step-by-step Ia

Centering



PCA Step-by-step II

Compute the covariance matrix

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top = \frac{1}{N}\underbrace{\bar{\mathbf{X}}\bar{\mathbf{X}}^\top}_{\text{Scatter Matrix } \mathbf{S}}.$$

Question: What is the difference between the covariance matrix of the original dataset $\mathbf{X}$ and that of the zero-mean data $\bar{\mathbf{X}}$?

PCA Step-by-step II

Eigenvalue decomposition: Compute the eigenvalue decomposition of the covariance matrix. Since $\mathbf{\Sigma}$ is symmetric,

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top,$$

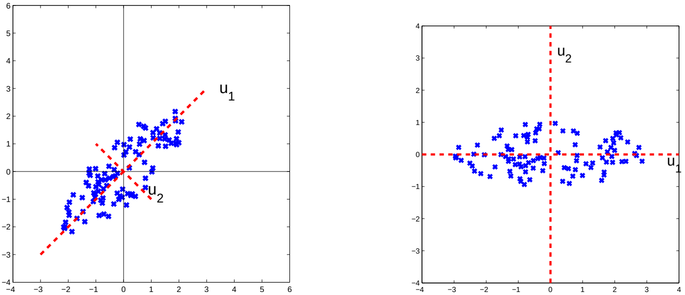
where $\mathbf{\Lambda} = \text{diag}[\lambda_1, ..., \lambda_D]$, such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_D$.

Eigenvector matrix $\mathbf{U} = [u_1, \dots, u_D]$, where $u_i \in \mathbb{R}^D$ are unit (i.e. $\|u_i\|_2 = 1$) and orthonormal column eigenvectors.

Question: How does the eigendecomposition of the scatter matrix $\mathbf{S}$ differ from that of $\mathbf{\Sigma}$?

PCA Step-by-step IIa

Eigenvalue decomposition and rotation



PCA Step-by-step III

Model selection: Pick a $K \leq D$ and keep the projections associated with the top K eigenvalues. (Capture maximal variance of the data.)

Transform the data onto the new basis of K dimensions:

Projection matrix: $\mathbf{U}_K = [u_1, \ldots, u_K] \in \mathbb{R}^{D \times K}$

$$\bar{\mathbf{Z}} = \mathbf{U}_K^\top \bar{\mathbf{X}}$$

$\bar{\mathbf{Z}} \in \mathbb{R}^{K \times N}$: We obtain a dimension reduction of the data.

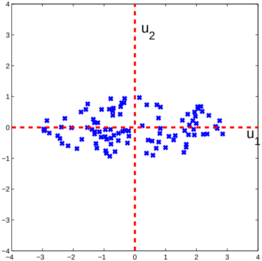
Reconstruction: Go back to original basis:

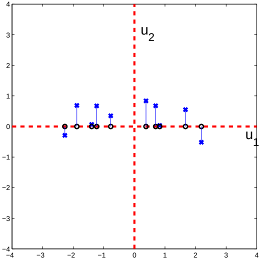
$$\bar{\tilde{\mathbf{X}}} = \mathbf{U}_K \bar{\mathbf{Z}}$$

and correct for shift $\tilde{\mathbf{X}} = \bar{\tilde{\mathbf{X}}} + \mathbf{M}$.

PCA Step-by-step IIIa

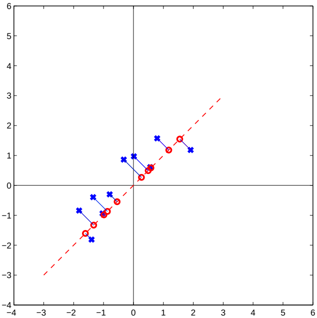
Scalar projection onto eigenvector subspaces

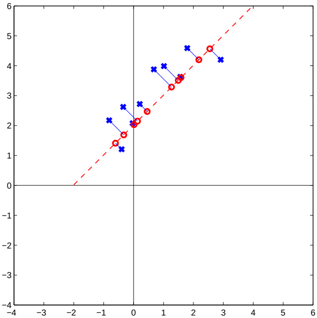




PCA Step-by-step IIIb

Inverse rotation and shift





PCA Reconstruction Error

$$\text{err} = \frac{1}{N}\sum_{i=1}^{N}\|\tilde{\bar{x}}_i - \bar{x}_i\|_2^2 = \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2$$

where $\|A\|_F = \sqrt{\sum_i \sum_j a_{ij}^2} = \sqrt{\text{trace}(AA^\top)}$ is the Frobenius norm of matrix A.

Goal:
Prove that

$$\text{err} = \sum_{i=K+1}^{D} \lambda_i$$

PCA Reconstruction Error - Proof

$$\text{err} = \frac{1}{N} \|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2$$

PCA Reconstruction Error - Proof

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

PCA Reconstruction Error - Proof

$$\begin{aligned} \mathsf{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\ &= \frac{1}{N}\mathsf{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \end{aligned}$$

PCA Reconstruction Error - Proof

$$
\begin{aligned}
\mathsf{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\
&= \frac{1}{N}\mathsf{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
&= \mathsf{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\boldsymbol{\Sigma}\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right)
\end{aligned}
$$

PCA Reconstruction Error - Proof

$$
\begin{aligned}
\text{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\
&= \frac{1}{N}\text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\boldsymbol{\Sigma}\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right)
\end{aligned}
$$

PCA Reconstruction Error - Proof

$$
\begin{aligned}
\text{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\
&= \frac{1}{N}\text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{\Sigma}\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top\mathbf{U} - \mathbf{U})\cdot\mathbf{\Lambda}\cdot(\mathbf{U}^\top\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{U}^\top)\right)
\end{aligned}
$$

PCA Reconstruction Error - Proof

$$
\begin{aligned}
\text{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\
&= \frac{1}{N}\text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{\Sigma}\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top\mathbf{U} - \mathbf{U})\cdot\mathbf{\Lambda}\cdot(\mathbf{U}^\top\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{U}^\top)\right) \\
&= \text{trace}\left(([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})\mathbf{\Lambda}([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})^\top\right)
\end{aligned}
$$

PCA Reconstruction Error - Proof

$$
\begin{aligned}
\text{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\
&= \frac{1}{N}\text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{\Sigma}\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top\mathbf{U} - \mathbf{U})\cdot\mathbf{\Lambda}\cdot(\mathbf{U}^\top\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{U}^\top)\right) \\
&= \text{trace}\left(([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})\mathbf{\Lambda}([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})^\top\right) \\
&= \text{trace}(\sum_{i=K+1}^{D}\lambda_i u_i u_i^\top)
\end{aligned}
$$

PCA Reconstruction Error - Proof

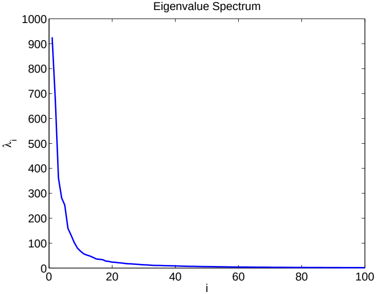
$$
\begin{aligned}
\text{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\
&= \frac{1}{N}\text{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{\Sigma}\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \text{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top\mathbf{U} - \mathbf{U})\cdot\mathbf{\Lambda}\cdot(\mathbf{U}^\top\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{U}^\top)\right) \\
&= \text{trace}\left(([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})\mathbf{\Lambda}([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})^\top\right) \\
&= \text{trace}(\sum_{i=K+1}^{D}\lambda_i u_i u_i^\top) = \sum_{i=K+1}^{D}\lambda_i\cdot\text{trace}\left(u_i u_i^\top\right)
\end{aligned}
$$

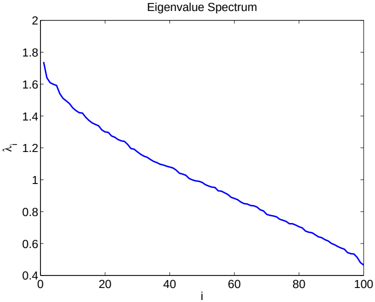
PCA Reconstruction Error - Proof

$$
\begin{aligned}
\mathsf{err} &= \frac{1}{N}\|\tilde{\bar{\mathbf{X}}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{N}\|(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\bar{\mathbf{X}}\|_F^2 \\
&= \frac{1}{N}\mathsf{trace}((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\bar{\mathbf{X}}\bar{\mathbf{X}}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)^\top) \\
&= \mathsf{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{\Sigma}\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \mathsf{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\cdot\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\cdot(\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{I}_d)\right) \\
&= \mathsf{trace}\left((\mathbf{U}_K\mathbf{U}_K^\top\mathbf{U} - \mathbf{U})\cdot\mathbf{\Lambda}\cdot(\mathbf{U}^\top\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{U}^\top)\right) \\
&= \mathsf{trace}\left(([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})\mathbf{\Lambda}([\mathbf{U}_K;\mathbf{0}] - \mathbf{U})^\top\right) \\
&= \mathsf{trace}(\sum_{i=K+1}^{D}\lambda_i u_i u_i^\top) = \sum_{i=K+1}^{D}\lambda_i\cdot\mathsf{trace}\left(u_i u_i^\top\right) \\
&\overset{unit\_norm}{=} \sum_{i=K+1}^{D}\lambda_i
\end{aligned}
$$

Reading the Eigenspectrum

Interpret eigenvalues as the variance in the dimension specified by the corresponding eigenvector.





For each eigenvalue spectrum, how many dimensions (K) should we keep?