
Artificial Intelligence and Machine Learning

Exercises – Principal Component Analysis

Question 1 (Eigendecomposition of a symmetric matrix)

Let the square (and symmetric) matrix

$$\mathbf{A} := \begin{pmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{pmatrix} \in \mathbb{R}^{3 \times 3}$$

be given. Please answer the following questions:

1. What are the eigenvalues of \mathbf{A} ?
2. Compute the eigenspace for each eigenvalue. What do you observe?
3. Write down the eigendecomposition of \mathbf{A} , i. e. find the matrices $\mathbf{U} \in \mathbb{R}^{3 \times 3}$ and $\mathbf{\Lambda} \in \mathbb{R}^{3 \times 3}$ such that $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$.

Question 2 ☹

Which statements concerning PCA are **false**?

- ☐ The algorithm belongs to the category of supervised learning.
- ☐ The goal of PCA is to reduce the variance in the data as much as possible.
- ☐ The data is transformed linearly.
- ☐ All principal components are orthogonal.

Question 3 (Covariance matrix) ☹

Let Σ be a covariance matrix. Which statements are correct?

- ☐ Σ is symmetric (i. e. $\Sigma^\top = \Sigma$).
- ☐ $\Sigma \in \mathbb{R}^{M \times N}, M \neq N$.
- ☐ Σ is a square matrix.
- ☐ All entries on the main diagonal of Σ are non-negative.

Question 4 (Number of principal components) ☒

Explain in up to three sentences how you can choose the number of principal components for dimensionality reduction. Name at least two different strategies!

Question 5 (PCA by hand) ☒

You are presented with a 4-dimensional dataset. Your task is to reduce the dimensionality of the dataset as much as possible, while preserving at least 80 % of the original variance.

You have already computed the covariance matrix Σ and its eigendecomposition $\Sigma = U\Lambda U^T$ in a previous step. You have found:

$$\underbrace{\begin{pmatrix} 4 & -2 & 3 & -1 \\ -2 & 4 & -3 & 1 \\ 3 & -3 & 5 & 0 \\ -1 & 1 & 0 & 5 \end{pmatrix}}_{\Sigma} \approx \underbrace{\begin{pmatrix} -0.5 & 0.7 & -0.5 & 0 \\ 0.5 & 0.7 & 0.5 & 0 \\ -0.6 & 0 & 0.7 & -0.3 \\ 0.2 & 0 & -0.2 & -1 \end{pmatrix}}_U \underbrace{\begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}}_{\Lambda} \underbrace{\begin{pmatrix} -0.5 & 0.5 & -0.6 & 0.2 \\ 0.7 & 0.7 & 0 & 0 \\ -0.5 & 0.5 & 0.7 & -0.2 \\ 0 & 0 & -0.3 & -1 \end{pmatrix}}_{U^T}$$

Let D be the dimensionality of the reduced dataset. Determine D such that the constraint above is satisfied. Which vectors represent the principal components? *Please justify your answer.*

**Question 6 (Eigenfaces)**

Sklearn provides the *Labeled Faces in the Wild (LFW)* people dataset which comprises images of well known people's faces. Each image has a resolution of 50×37 pixels. The data can be loaded using the following commands:

```
1 from sklearn.datasets import fetch_lfw_people

3 # GET DATA
  # -----
5 lfw_people = fetch_lfw_people(
    min_faces_per_person=70, resize=0.4)

7

9 # get number of samples, height, and width of images
  n_samples, h, w = lfw_people.images.shape

11 # get features
    X = lfw_people.data
```

Remark: Each image is labeled with the name of the person shown. The labels can be loaded with the command `y = lfw_people.images.target`. However, we are not going to use these in this exercise. Figure 1 shows some sample images.

Work through the following tasks:

- Compute the mean face and the eigenfaces by performing an eigendecomposition of the covariance matrix of the data. For this you may want to use the function `np.linalg.eig()`.

- Plot the first thirty eigenfaces.



Figure 1: Some sample images from the LFW dataset.