Prof. Dr. Jan Rolfes, Ehsan Waiezi, Lars Weidner
Winter term 23/24

Mathematics of Learning – Worksheet 3

- The exercise sheets will be uploaded every Monday. Solution sketches will be uploaded one week later.
- You can hand in your own solutions via StudOn and we correct them - this is not mandatory. Please hand in in small groups of 2-3 students.
- For questions, please either use the forum on StudOn or send an email to jan.jk.krause@fau.de if it is a personal question.

Basics [Analytically calculating eigenvalues and eigenvectors].[1]

Let $A \in \mathbb{R}^{n\times n}$ be a quadratic matrix. Whenever for a vector $v \in \mathbb{R}^n$ and for a $\lambda \in \mathbb{R}$ the equation

$$Av = \lambda v$$

holds, we call λ an eigenvalue of A and v the corresponding eigenvector. Find out how to calculate eigenvalues and eigenvectors analytically and calculate the eigenvalues and eigenvectors of the matrix

$$A := \frac{1}{3} \cdot \begin{pmatrix} 5 & -2 & -1 \\ -2 & 5 & 1 \\ -1 & 1 & 8 \end{pmatrix}.$$

Hint: The eigenvalues of this matrix are integers; if you get some fractional values, you made some mistake.

Exercise 1 [Definiteness of the sample covariance matrix].
Let $y^{(1)}, \dots, y^{(N)} \in \mathbb{R}^M$ be zero-centered input data and let C be the respective sample covariance matrix.

1. Show that C is always positive semi-definite.
2. In which cases is $\langle x, Cx\rangle = 0$ for an $x \in \mathbb{R}^M \setminus \{\vec{0}\}$. What does that mean for the given data?

Exercise 2 [Prerequisites for PCA].
Given a set of data vectors $x_1, ..., x_N \in \mathbb{R}^p$ and a matrix $V \in \mathbb{R}^{p\times q}$, $q < p$, with q orthogonal unit vectors as columns. Prove that

$$\tilde{\mu} = \bar{x}, \quad \tilde{\lambda}_i = V^T(x_i - \bar{x})$$

[1] There are lots of nice tutorial books for linear algebra and analysis available in our library, one of them I put in the forum. For a less formal introduction, you can also consult wikipedia (with caution).

is a minimizer (over μ and λ_i) of

$$f(\mu, \lambda_1, ..., \lambda_N) = \sum_{i=1}^{N} ||x_i - \mu - V\lambda_i||^2,$$

where $||\cdot||$ denotes the euclidean norm. Furthermore, show that the minimizer $\bar{x}$ for μ is not unique and find the set of minimizers for μ.

Exercise 3 [Implementing PCA for data reduction].
Implement the (linear) principal component analysis algorithm as described on the slides. For the numerical approximation of the eigenvalues and respective eigenvectors of the covariance matrix C you can use the Python function `scipy.linalg.eig`.
Test your algorithm on the Iris data set[2]. This is perhaps the most popular data set to be found in the pattern recognition literature. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. The data has 5 columns representing the following attributes:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:

- Iris Setosa
- Iris Versicolor
- Iris Virginica

You can ignore attribute 5 in the above list for the PCA, but use it for visualization of the different classes. Plot the features after applying the PCA algorithm for $k = 3$ and $k = 2$. What can you observe?

Exercise 4 [Apply clustering algorithms].
Apply the clustering algorithms (k-means and EM) to the Iris data set (before and after PCA). Describe, interpret and visualize your results.

[2]Fisher,R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936);