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

Mathematics of Learning – Worksheet 4

- 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 use the forum on StudOn since other students may have similar questions. If you have a more personal question about the exercises please send an email to ehsan.waiezi@fau.de or lars.weidner@fau.de respectively.

Basics [Eigenvectors of symmetric matrices.] Consider a symmetric matrix $A \in \mathbb{R}^{n\times n}$ (i.e. $A = A^T$). Prove that for two eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ with $\lambda_1 \neq \lambda_2$ and corresponding eigenvectors $v_1 \in \mathbb{R}^n$ and $v_2 \in \mathbb{R}^n$ it holds that $\langle v_1, v_2 \rangle = 0$, i.e., that v_1 and v_2 are orthogonal.

Basics. [Singular Value Decomposition].
Calculate the SVD of the matrix

$$A = \begin{pmatrix} 2 & -2 & -2 & 0 \\ -1 & -1 & 3 & 4 \\ 2 & -2 & 2 & -2 \end{pmatrix}.$$

The SVD of a matrix $A \in \mathbb{R}^{m\times n}$ consists of matrices $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$ which are orthogonal, i.e., UU^T, VV^T is equal to the corresponding unit matrix, and a matrix $\Sigma \in \mathbb{R}^{m\times n}$ which has only positive, descending entries on the diagonal (it looks like

$$\begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & 0 \end{pmatrix}$$

in our case). $U \cdot \Sigma \cdot V^T = A$ should hold, in case you calculated correctly.
Hint: No nice numbers this time.

Exercise 1 [Reading assignment: Association rules].
Read chapter 14.2 of the *Hastie* book regarding association rules. Discuss the contents with one (or more) fellow student for at least half an hour.

Exercise 2 [Reading assignment: Self organizing maps (SOM)].
Read chapter 14.5 of the *Hastie* book (or 14.4, depending on your version) regarding self organizing maps. Discuss the contents with one (or more) fellow student for at least half an hour.

Exercise 3 [Equivalence of eigenvalue problems].
Let $x^{(1)}, \ldots, x^{(N)}$ be given input data. Furhermore, let $\mathcal{H}$ be a (possibly infinite-dimensional) Hilbert space, $\Psi \colon \mathbb{R}^M \to \mathcal{H}$ a map from the input data, $\mathbf{C}$ the covariance matrix of the transformed data in $\mathcal{H}$ with:

$$\mathbf{C} := \frac{1}{N}\sum_{i=1}^{N} \Psi(x^{(i)})\Psi(x^{(i)})^T, \tag{1}$$

Furthermore, let $k \colon \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ be the corresponding kernel function and K the associated Kernel matrix with $K_{i,j} = k(x^{(i)}, x^{(j)})$.

1. Show that for any $\lambda \neq 0$ every solution $\vec{\alpha} \in \mathbb{R}^N$ with $\vec{\alpha} \perp \mathrm{kern}(K)$ of the equation

   $$N\lambda K\vec{\alpha} = K^2\vec{\alpha} \tag{2}$$

   is also a solution of the eigenvalue equation:

   $$N\lambda\vec{\alpha} = K\vec{\alpha}. \tag{3}$$

2. Use the previous statement to show that the following "equivalence" holds for all $\lambda > 0$
   (a) $\mathbf{v} \in \mathcal{H}$ is eigenvector of $\mathbf{C}$ with respect to eigenvalue $\lambda \Rightarrow$
   $\vec{\alpha} \in \mathbb{R}^N$ defined such that $\mathbf{v} = \sum_{i=1}^{N} \alpha_i \Psi(x^{(i)})$ is eigenvector of K with respect to eigenvalue $N\lambda$
   (b) $\vec{\alpha} \in \mathbb{R}^N$ is eigenvector of K with respect to eigenvalue $N\lambda \Rightarrow$
   $\mathbf{v} = \sum_{i=1}^{N} \alpha_i \Psi(x^{(i)}) \in \mathcal{H}$ is eigenvector of $\mathbf{C}$ with respect to eigenvalue λ

Exercise 4 [Implementing Kernel PCA for data reduction].
Implement the Kernel principal component analysis algorithm as described on the slides. For the numerical approximation of the eigenvalues and respective eigenvectors of the Kernel matrix K you can use the Python function `scipy.linalg.eig`.
Test your algorithm on the "Circle" data set. Each line of the data file has to be interpreted as a single data point with `[x, y, label]`. Compare the effect of using an inhomogeneous polynomial kernel of degree 2 and a Gaussian kernel by plotting the respective first two principal components. Choose a good value for $\sigma^2 > 0$ and $a \in \mathbb{R}$ in case of the Gaussian kernel and the polynomial kernel, respectively.