Florian Rösel, Prof. Dr. Frauke Liers Winter semester 2021/2022

Mathematics of Learning – Worksheet 4

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

Exercise 1 [Reading assignment: Spectral Clustering].
Read chapter 14.5.3 regarding Spectral Clustering in the Hastie book. Spectral Clustering is a method which can be applied to data with some radial structure, for example. At some point in the chapter, the Laplacian of graphs will be of importance. If you do not know about graph Laplacians, inform yourself about it (it is going to be important later in the lecture). Peculiarly ambitious students can implement their version of Spectral Clustering and apply it on various data sets (extract some data from the internet or use the data sets already uploaded or which you generated on your own - be creative). Discuss the contents of the chapter with a fellow student for at least half an hour.

Leaving

Exercise 2 [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:

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

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

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

   is also a solution of the eigenvalue equation:

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

2. Use the previous statement to show that the following equivalence holds for all $\lambda > 0$:

$$\begin{gathered} \mathbf{v} \in \mathcal{H} \text{ is eigenvector of } \mathbf{C} \text{ with respect to eigenvalue } \lambda \\ \Leftrightarrow \\ \vec{\alpha} \in \mathbb{R}^m \text{ is eigenvector of } K \text{ with respect to eigenvalue } N\lambda \end{gathered}$$

Exercise 3 [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.

Acknowledgement, for generating a previous version of this sheet, to Leon Bungert and Daniel Tenbrinck.