Assignment 10

Machine Learning: Algorithms and Theory

Prof. Ulrike von Luxburg / Diego Fioravanti / Moritz Haas
Tobias Frangen / Siavash Haghiri

Summer term 2018 — due to **July 3**

1. When you hand in your assignment you need to hand in the notebook too. Please do not write down a report with just results and/or figures. Ideally you should email it, it is easier to correct and more eco friendly, but we accept printed versions too. From now on not handing in the notebook will result in 0 points for the programming part.

2. Before the end of the course you need to present at least one of your solution in the tutorial. If you do not do that you cannot take the exam! If for any reason you cannot attend let us know, it is possible to change group or find another solution.

3. Join the class on ILIAS otherwise we cannot contact you if we need to.

Exercise 1 (Spectral graph theory 3+1+3+1 points) In the following we consider the symmetric normalized graph Laplacian $L = D^{-1/2}(D - A)D^{-1/2}$, where A denotes the adjacency matrix of an undirected and unweighted graph and D the graph's degree matrix, for various special graphs.

(a) Let G be an undirected, unweighted and connected graph. Let $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n$ denote the increasingly sorted eigenvalues of L Prove that

$$\lambda_n \leq 2.$$

(b) Let G be the complete graph on n vertices, that is any two vertices are connected. Show that its Laplacian has eigenvalues 0 with multiplicity one and $n/(n-1)$ with multiplicity $n-1$.

(c) Let G be the complete bipartite graph on $n+m$ vertices, that is every vertex $v \in \{v_1, \ldots, v_n\}$ is connected with every vertex $\tilde{v} \in \{v_{n+1}, \ldots, v_{n+m}\}$ but no two vertices in $\{v_1, \ldots, v_n\}$ nor $\{v_{n+1}, \ldots, v_{n+m}\}$ are connected. Show that its Laplacian has eigenvalues 0 with multiplicity one, 1 with multiplicity $m+n-2$ and 2 with multiplicity one.

(d) Let G be the star on n vertices, that is v_1 is connected with every vertex $v \in \{v_2, \ldots, v_n\}$ but no two vertices in $\{v_2, \ldots, v_n\}$ are connected. Show that its Laplacian has eigenvalues 0 with multiplicity one, 1 with multiplicity $n-2$ and 2 with multiplicity one.

Hint: One way to construct the eigenvectors is to build the characteristic polynomial via induciton.

Exercise 2 (Implementing k-means and spectral clustering, 3+1+4+1 points)

Use jupyter notebook for this exercise, not jupyter-lab.

(a) Implement k-means in the function `KMeans(X, k)` which takes as input data points X_i with $i = 1, \ldots, n$ and parameter k. It computes the clustering and returns a vector of length n in which the i-th entry is from $\{1, 2, \ldots, k\}$ representing the cluster membership of the i-th data point. It also returns the cluster centers.

Points for this task will be granted for a correct implementation. In the next tasks you can use the built-in function `sklearn.cluster.KMeans` if you are not confident in your own implementation.

(b) In X you find a $n \times 2$ matrix with $n = 400$ samples in $\mathbb{R}^2$ from three clusters. Apply k-means with $k = 3$, plot the data and color the points according to their assigned cluster. Mark the cluster centers. Repeat this for $k = 2$ and $k = 4$.

(c) Implement a function `generateKNNgraph(X,k)` which takes as input the data points (X_i) with $i = 1, \ldots, n$ and parameter k. Return the $n \times n$ weight matrix of the corresponding kNN-graph W where $W_{i,j} = 1$ if X_i is among the k nearest neighbors of X_j or vice versa. Otherwise set $W_{i,j} = 0$.

Implement a function `spectralClustering(knnW,k,normalize)` which takes as input a kNN-graph, parameter k for the number of clusters (not for the kNN-graph) and a boolean value indicating whether to use the normalized or unnormalized Laplacian. Perform spectral clustering, run k-means on the spectral embeddings and return the the centers, the clustering as well as the embedded data points as a $n \times k$ matrix.

(d) Apply unnormalized spectral clustering to the data set from task (b). Choose a reasonable parameter k for the kNN-graph and try to detect three clusters. For the resulting clustering make the same plot as in (b) to visualize the cluster membership of each point. Repeat this with normalized spectral clustering. Visualize the spectral embedding in a 3-d scatter plot.

Exercise 3 (Design your own exam questions, 3 points) In this exercise, everybody is supposed to come up with suggestions for three exam questions. This is a good way to recap/understand the concepts discussed so far.

Put yourself in our place! We do not want to ask stupid questions. We would like to ask "nice questions". In general, written exams contain three types of questions:

- Questions that are just about **reproducing** knowledge. These kind of questions are pointless and we don't ask questions of this kind in the exam.
- Questions for testing whether the person **understands** the concepts and can apply them to simple situations.
- Questions that require to **transfer** knowledge to new situations.

Your task is now to design exam questions along with their solutions of the two last-mentioned types. Enter your questions in the LaTex file `my_exam_questions.tex` that we provided and send it to your tutors.

After the class we will put all your questions online. At the end of the course, these questions can help everybody to prepare for the exam!