9 Hierarchical and Spectral Clustering

You can use external libraries for linear algebra operations but you are expected to write your own algorithms.

9.1 Exercise 1

Use the `iris` dataset available at `https://archive.ics.uci.edu/dataset/53/iris`.

- Apply **your own** PCA implementation to the dataset. Plot the data in two dimensions, coloring by class.
- Perform hierarchical clustering with single linkage and ward's linkage using `scipy` functions (see `https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html`).
- Plot the dendogram and the datapoints in 2D, coloring now by cluster (cut the dendogram to have the same number of clusters as the ground truth).
- Implement your own version of spectral clustering.
- Apply it to the dataset, setting k to be the same of the ground truth (build the graph with k-NN having $k = 5$). Plot the datapoints in 2D, coloring by cluster.

Notes

Spectral Clustering

For Spectral Clustering, take into account that k-NN is not a symmetric relationship but the Graph you build from it will be symmetric. Take care of specifying all the edges between couple of points!

The weights of the edges are not the distances, but the similarities! A naive approach would be setting $s_{ij} = 1$ if there is an edge, and $s_{ij} = 0$ if there is none. A better approach would be to use a decreasing function of the distances.

Once you have a working implementation of Spectral Clustering, feel free to investigate its different versions, as discussed in class.

PCA projections

We use PCA only to plot the data in 2 dimensions. You will have to use the original dataset to compute the clusters, and then use the obtained results to color the PCA projections.