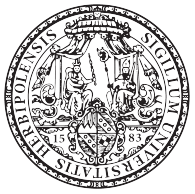

Introduction to Informatics
WiSe 2024/2025

Prof. Dr. Ingo Scholtes
Chair of Informatics XV
University of Würzburg

Exercise Sheet 07

Published: December 19, 2024
Due: January 07, 2024

Please upload your solutions to WueCampus as a scanned document (image format or pdf), a PDF document, and/or as a `jupyter` notebook.

1. **K-Means**

   The elbow method is a technique used to determine the optimal number of clusters k in a clustering problem, particularly for algorithms like K-means. It involves evaluating a loss function for different values of k and identifying a point where adding more clusters no longer significantly improves the clustering results.

   Use the provided notebook to

   (a) apply `k-means` to cluster the blob dataset for diffenrent values of k. Use $k = 1, ..., 10$.

   (b) Compute the given loss function for each k

   (c) plot the loss values against the number of clusters k

   (d) Analyze the plot to indentify the "ellbow point", where the loss function stops decreasing significantly

   (e) Identify the optimal number of clusters k based on the elbow point

2. **Multi Layer Perceptron**

   In last week's exercise, we used **Logistic Regression** to classify the moons and circles datasets. However, as we observed, a linear model like Logistic Regression is not powerful enough to accurately separate the classes in these non-linear datasets.

   This week, we will address this limitation by using a **Multi-Layer Perceptron (MLP)** model. The MLP, being a neural network, can capture non-linear relationships, making it well-suited for this task.

   a) Use the provided notebook to train an `MLPClassifier` on both the moons and circles datasets.

   b) Experiment with the `hidden_layer_sizes` parameter, which defines the number of neurons in each hidden layer. For example:

      - `(10,)` represents a single hidden layer with 10 neurons.
      - `(50, 30)` represents two hidden layers with 50 and 30 neurons, respectively.

   c) Evaluate how different numbers of layers and neurons impact the model's performance

   d) Compare the MLP model's performance with the Logistic Regression model from last week.