Aprenentatge Automàtic per a Xarxes (ML4Net)

Seminar 3 - Problem statement

May 2, 2025

Abstract

On this occasion, we will focus on Wi-Fi fingerprinting, which is a low-complexity technique to locate users within a given area. You will derive a *k-means* model that predicts the location of users based on their Received Signal Strength Indicator (RSSI).

Disclaimer: The data used in this seminar were synthetically generated, but the aim is to reproduce real-world measurements.

1 Part I: Data preparation & analysis

The dataset for this seminar (`dataset_Seminar3.zip`) contains two files:

1. `rssi_data.csv`: Wi-Fi RSSI measurements taken at $K = 3$ different APs (AP1, AP2, and AP3), for $P = 5$ positions in which the STA was placed. In each row, there is the RSSI perceived by each of the APs for a single measurement. The dataset includes $M = 300$ measurements per location $p \in P$, thus it contains $(P \times M) \times K$ values.

2. `labels_data.csv`: The ground truth (i.e., the real position of the STA) for each measurement. $(P \times M)$ values are provided.

Once you load the data into your workspace, process it to provide the following plots:

- 2-D plots showing the RSSI perceived by each pair of APs (e.g., AP1 vs AP2).
- A 3-D plot showing the RSSI perceived by each of the APs.

Apart from that, describe the statistics behind each of the considered points. For that, compute the average RSSI and the standard deviation obtained in each of the 5 studied locations. Plot the results.

2 Part II: k-means implementation

Implement a simple k-means algorithm (by hand) that performs the following operations:

- Initializes the centroids randomly in 3D space.

- Iteratively updates the centroids by following these two steps:
  - Assign each point in the dataset to a centroid by calculating the Euclidean distance between the point and each centroid. The selected centroid should be the one that is closest in distance to the point.
  - Update the centroid of each cluster as the mean position of the points in that cluster.[1]
- Check if the algorithm has converged. For that, establish two conditions:
  - The assignment of clusters to each point in the dataset does not vary from one iteration to another.
  - A maximum number of iterations ($T = 100$) is reached.

Once your k-means implementation is ready, use it in the Wi-Fi fingerprinting dataset to determine the location of each point. Compare the results with the ground truth and plot the results.

[1] In case a cluster is empty, you will have to re-initialize the corresponding centroid randomly.