DS502 Project Prosal

- members names: Xiuhan LI, Enbo Tian, Ziyang Xu
- description of the problem

  Many real world applications need to know the localization of a user in the world to provide their services. Therefore, automatic user localization has been a hot research topic in the last few years. Automatic user localizatio n consists of estimating the position of the user (latitude, longitude and altitude) by using an electronic device, usually a mobile phone.

  Our project is focus on the problem of indoor localization. We are going to estimate the local place of users, using the dataset of 250 WAPs signal str ength. We will decide the building and floor of the user from machine learn ing.

- description of the dataset

  The dataset we used is from UCI machine learning website (https://archive.i cs.uci.edu/ml/datasets/ujiindoorloc). This dataset covers three buildings o f Jaume I University with 4 or more floors and almost 110.000m2. The 529 at tributes contain the 520 Wi-Fi fingerprints, longitude, latitude, floor, bu ilding, space id, relative position, user id, phone id, timestamp. Each Wi-Fi fingerprint can be characterized by the detected Wireless Access Points (WAPs) and the corresponding Received Signal Strength Intensity (RSSI). The intensity values are represented as negative integer values in the ranging of -104dBm (extremely poor signal) to 0dbM. The positive value 100 represen ts no WAP signal detected. During the database creation, 520 different WAPs were detected. Thus, the Wi-Fi fingerprint is composed of 520 intensity val ues.

  Attribute Information:

  - Attribute 001 (WAP001): Intensity value for WAP001. Negative integer values from -104 to 0 and +100. Positive value 100 used if WAP001 wa s not detected.
  - Attribute 520 (WAP520): Intensity value for WAP520. Negative integer values from -104 to 0 and +100. Positive Value 100 used if WAP520 wa s not detected.
  - Attribute 521 (Longitude): Longitude. Negative real values from -769 5.9387549299299000 to -7299.786516730871000
  - Attribute 522 (Latitude): Latitude. Positive real values from 48647 45.7450159714 to 4865017.3646842018.
  - Attribute 523 (Floor): Altitude in floors inside the building. Integ er values from 0 to 4.

    - Attribute 524 (BuildingID): ID to identify the building. Measures were taken in three different buildings. Categorical integer values from 0 to 2.
    - Attribute 525 (SpaceID): Internal ID number to identify the Space (office, corridor, classroom) where the capture was taken. Categorical integer values.
    - Attribute 526 (RelativePosition): Relative position with respect to the Space (1 - Inside, 2 - Outside in Front of the door). Categorical integer values.
    - Attribute 527 (UserID): User identifier (see below). Categorical integeger values.
    - Attribute 528 (PhoneID): Android device identifier (see below). Categorical integer values.
    - Attribute 529 (Timestamp): UNIX Time when the capture was taken. Integer value.
- classification or regression: This is a classification problem
- the methods you plan to try: We plan to try logistic regression (with softmax regression) and decision tree to predict the labels.
- the error metrics you plan to use and the algorithms for assessing them: Since the error terms will follow iid Normal distributions, where the error terms are homoscedastic, uncorrelated. Then we can use heteroskedasticity and
  autocorrelation consistent estimator std errors.
- comments and concerns：Since we want to predict two labels, building and floor. But the algorithm that we choose can only predict one label at a time. So we decided to try two ways: one is to combine and embed building and floor into a single value label, another is to train models on building and floor separately. And finally evaluate the accuracy.