ICT for Smart Mobility

Exercise – Transport modelling - 2nd part
Creating trip generation model of demand with KDE

In this second part of the practical activity, you will create a demand model for trip generation in continuous space.

With KDE, you do NOT need to choose an appropriate zoning system. The demand model should generalize and you will perform calibration and validation process.

Data

Same dataset "Amsterdam_dataset.csv" as in the previous exercise

STEP A – KDE for trip generation

- Implement a simple KDE with Gaussian Kernel and arbitrary h or use the KernelDensity model from sklearn https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KernelDensity.html
- Fit this model to the data
- Obtain the model for P(O) and P(D)
- Show the distributions P(O) and P(D) of the trip generation model on a map

STEP B – Calibration of trip generation – finding optimal bandwidth h

- Evaluate the total log likelihood metric to be maximized: 1/N sum_(samples s) log(P(s)). You can also use score_samples of KernelDensity class of sklearn
- Calibrate the bandwidth h in order to maximize loglikelihood (go back to STEP A and recompute the trip generation model with different h)
- What is a good value for h? What is the corresponding loglikellihood?

STEP C – Validation of trip generation – finding bandwidth h

- Divide the sample dataset into random training and validation samples and fit the KDE models (origin and destination) to the training data. You can also use k-fold cross validation with a small k (e.g., k=5).
- Evaluate the total log likelihood of the training and validation
- Calibrate the bandwidth h as in STEP B, but now on the TRAINING in order to maximize loglikelihood on the VALIDATION
- What are the differences with respect to step B?
- Show the distributions P(O) and P(D) of the trip generation model on a map