

Aflevering 1: Clustering

Sune Lauth Gadegaard

Efterår 2022

I denne opgave tages udgangspunkt i datasættet **USArrests.csv**. Heri er angivet statistik for antallet af arrestationer per 100.000 indbyggere for overfald, mord og voldtægt i hver af de 50 stater i USA i 1973. Vi vil her antage, at de amerikanske stater skal inddeles i 5 grupper baseret på disse data.

Det er nu op til jer at besvare de følgende opgaver:

- Opgave 1:** Vælg mindst én metrik (afstandsmål), som skal bruges til at måle hvor ens to stater er (I kan blandt andet se på Euklidisk afstand og/eller Taxi-cap normen). Brug metrikken/metrikkerne til at udregne afstandsmatricen/afstandsmatricerne c_{ij} (se evt. afsnittet om lokationsbaseret clustering i noterne).
- Opgave 2:** Opstil det lokationsbaserede IP med sum-objektfunktion fra Clustering-noten for datasættet **USArrests.csv**. Implementer modellen i Python og løs det vha. en solver.
- Opgave 3:** Visualiser løsning. Dette kan for eksempel gøres ved at farve staterne i en cluster samme farve (se evt. [online værktøjet her](#) eller lignende).
- Opgave 4:** Opstil nu min-max udgaven af det lokationsbaserede IP fra noterne omkring Clustering og løs det tilhørende program. Visualiser ligeledes denne løsning.
- Opgave 5:** Opstil nu en model, der minimerer der største diameter over alle clustre. Implementer modellen i Pyomo og løs vha. en solver. Visualiser din løsning
- Opgave 6:** Kommenter på ligheder og uoverensstemmelser mellem de tre foregående løsninger. Hvad er forskellene og hvorfor ser vi disse forskelle (vi er ikke ude efter en socioøkonomisk fortolkning, men nærmere en forklaring på hvorfor ændringen af modellen ledte til disse nye løsning.)?

Jeres besvarelse af ovenstående opgaver skal sammenfattes i en lille rapport, som uploades i Peergrade inden afleveringsfristen. Sørg for at være grundige i jeres beskrivelser og analyser.