IIND4132

Homework 1

Instructor: Andrés Gómez, gomezand@usc.edu

1 Unsupervised learning

Consider the datasets depicted in Figure 1 (the data is provided in supporting files). The goal is to partition the datasets into three clusters, such that each cluster contains points with similar characteristics. In the HW we seek to formulate mixed-integer (possibly nonlinear) optimization problems that partition the points optimality according to a given criterion.

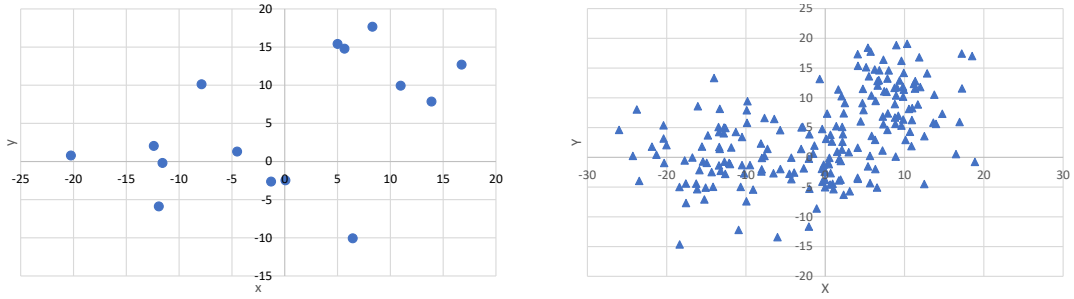

Figure 1: Small (left) and large (right) datasets for clustering

1. **Facility location** The goal is to select three points from the dataset to serve as centroids, such that the sum of the distances from each point to its closest centroid is minimized.
   a) Formulate this clustering problem as a mixed-integer optimization problem. Clearly define your variables, constraints and notation used.
   b) Implement the formulation in the language of your choice, and solve it with the small dataset. What is the optimal solution?
   c) Solve it with the large dataset (if necessary, use the NEOS server). What is the optimal solution?

2. **K-means** The goal is to choose three centroids (which may not correspond to points in the original dataset), such that the sum of the squared distances from each point to its closest centroid is minimized.
   a) Formulate this clustering problem as a mixed-integer optimization problem. Clearly define your variables, constraints and notation used.
   b) Implement the formulation in the language of your choice, and solve it with the small dataset. What is the optimal solution?
   c) (Optional) Solve it with the large dataset. What is the optimal solution/ best solution found?

3. **K-center** The goal is to select three points from the dataset to serve as centroids, such that the maximum distance from a point to its closed centroid is minimized.
   a) Formulate this clustering problem as a mixed-integer optimization problem. Clearly define your variables, constraints and notation used.
   b) Implement the formulation in the language of your choice, and solve it with the small dataset. What is the optimal solution?
   c) (Optional) Solve it with the large dataset. What is the optimal solution/ best solution found?

2 Portfolio optimization

Consider the mean-variance optimization problem seen in class (June 20). Given a set $\{1, \ldots, n\}$ of assets, letting x_i be the percentage of funds invested in asset i, and letting $z_i = 1$ if and only if asset i is included in the portfolio, the problem can be formulated as

$$\begin{aligned}
\min \ & \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j && && \text{(variance)} \\
\text{s.t.} \ & \sum_{i=1}^{n} x_i = 1 && && \text{(budget)} \\
& \sum_{i=1}^{n} \mu_i x_i \geq \beta && && \text{(target return)} \\
& \ell_i z_i \leq x_i \leq u_i z_i && i = 1, \ldots, n && \text{(min and max buy-in levels)} \\
& \boldsymbol{x} \in \mathbb{R}^n, \ \boldsymbol{z} \in \{0,1\}^n, &&
\end{aligned}$$

where: • σ_{ii} is the variance of asset i, and σ_{ij} is the covariance between assets i and j; • μ_i is the expected return of asset i; • β is the target return of the portfolio; • ℓ_i and u_i are the minimum and maximum transaction levels of asset i, if a transaction occurs at all.

Modify the formulation above to account for the following considerations (each consideration is independent from the others):

1. The portfolio must have at most k assets.
2. There is a fixed cost c_i incurred when investing any money in asset i. Ensure that the target return is satisfied after accounting for the fixed costs.
3. If asset 1 is included in the portfolio, then asset 2 must be included as well.
4. If asset 3 is included in the portfolio, then asset 4 cannot be included.
5. Assets with even indexes must account for at least half of the total budget.