# ALGORITHMS FOR OPTIMIZATION AND INFERENCE - 2024
# First Assignment

Matilde Dolfato

October 2024

## <u>Problem 2:</u>

Write an IP model for the k-means problem.

## Solution:

The k-means problem can be formulated as follows. The inputs are:

- a set $\boldsymbol{S}$ of colors, which contains all the colors of the image $s_i \in \mathcal{R}^3$, $i = 1, \dots, n$;
- a set $\boldsymbol{C}$ of centers, which contains all the possible centers (centroids) of the clusters $c_j \in \mathcal{R}^3$, $j = 1, \dots, m$;
- an integer $\boldsymbol{k}$, the desired number of clusters;
- a set $\boldsymbol{d}$ of distances, which contains the dissimilarity attained by each possible cluster, which is mapped to the set $C$;
- a matrix $\boldsymbol{z}$ of assignments, with entries $z_{ij} = 1$ if color $x_i$ is included in the cluster associated with center $c_j$ and 0 otherwise, with $i = 1, \dots, n$ and $j = 1, \dots, m$.

Notice that we could either consider the set $S$ of colors or the total set of pixels (which is the same, with repetitions). In this solution, we choose to consider only the set $S$, and indeed distances and centroids are computed weighting each observation in $S$ for the number of pixels of that color.

We introduce a binary decision variable $\boldsymbol{y_j}$, $y_j = 1$ if center $c_j$ is selected and 0 otherwise, $j = 1, \dots, m$.

Then, we consider the following objective function and constraints:

$$
\begin{aligned}
\min_{y_j} \quad & \sum_{j=1}^{m} y_j d_j \\
\text{s.t.} \quad & \sum_{j=1}^{m} y_j = k \\
& \sum_{j=1}^{m} z_{ij} y_j = 1 \qquad \forall i = 1, \dots, n \\
& y_j \in \{0, 1\} \qquad \forall j = 1, \dots, m
\end{aligned}
\tag{1}
$$

Here, we minimize the sum of dissimilarities attained by each selected center (selection is controlled by $y_j$), thus the total dissimilarity in our chosen clusters. The first constraint checks that exactly $k$ centers are chosen; the second that each color is assigned to exactly one center (or cluster). Indeed, we can think of the second constraint as checking that in the matrix $z$, each row contains exactly one 1 entry after multiplying by $y_j$. Lastly, we add the integer constraint to $y_j$, in particular that it is binary.