

# K-means Algorithm

Jacob Goldberger

Input data:  $x_1, \dots, x_n \in R^d$

Output centres:  $c_1, \dots, c_k \in R^d$

K-means algorithm:

- Assignment step:

$$c(x_t) = \arg \min_k \|x_t - c_k\|, \quad t = 1, \dots, n$$

- Update step:

$$c_k = \frac{\sum_{\{t|c(x_t)=k\}} x_t}{|\{t|c(x_t) = k\}|}, \quad k = 1, \dots, K$$

Cost function (that is monotonically decreasing):

$$C(c_1, \dots, c_k) = \sum_t \min_k \|x_t - c_k\|^2$$