CS 480/680 Homework 7
Out March 18, 2026
Due March 30, 2026

©Marina Meilă and Benjamin Schneider
mmp@uwaterloo.ca

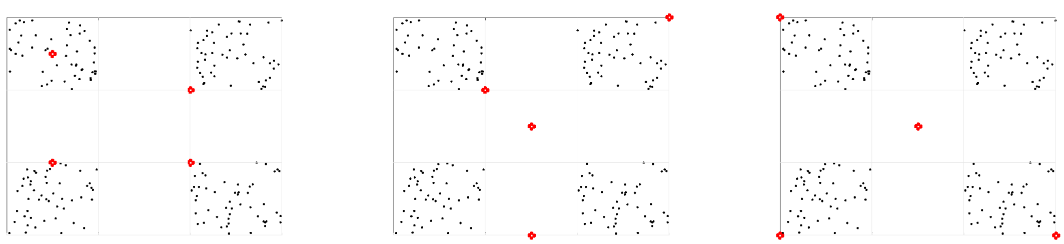
Problem 1 – Kmeans++ and K-logK

The data $\mathcal{D} = \{x^1, \ldots x^n\}$. You are initializing K-means with K clusters by different methods. In each case, calculate the number of distance computations necessary for the initialization as functions of K and n; you can ignore constant terms but not logarithmic or lower order terms.

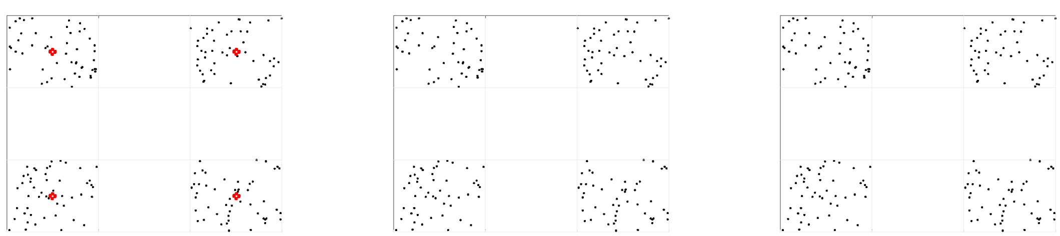
1.a Sample K centers uniformly at random without replacement from $\mathcal{D}$.
1.b Choose K centers from $\mathcal{D}$ by Fastest First Traversal.
1.c Choose K centers by Kmeans++.
1.d Choose the K centers by the K-logK method, with $K' = \texttt{round}(K \ln K)$ (assume that $K \ln K$ is an integer in your evaluation).
1.e Order the four methods above by the number of distance computations, assuming $n \gg K$ and $n - K \approx n$.

Problem 2 – Simple exercise with K-means

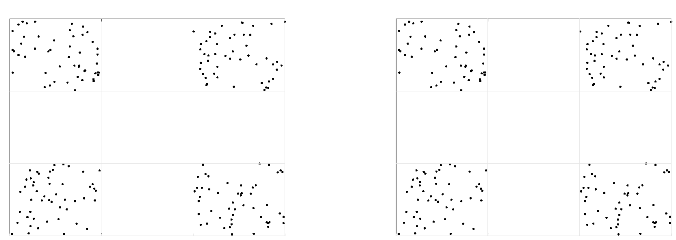
Example – initial centers | **a.** Initial centers | **b.** Initial centers



– after 1 step | – after 1 step | – after 1 step



final | final



Draw the position of the centers after 1 step of K-means in the cases **a, b** above.

Problem 3 – K-means facts

a. Prove that the K-means loss can be written as the sum of the squared distances between all pairs of points in the same cluster.

$$\mathcal{L}(\Delta, \mathcal{D}) = \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,j \in C_k} \|x^i - x^j\|^2 \tag{1}$$

b. CS 680 students only Assume that $n = mK$ and that the data set contains K equal clusters. You initialize by picking K data points at random, and assigning them to centers $\mu_{1:K}$. For simplicity, we write this as $\mu_{1:K} \sim uniform(\mathcal{D})$. What is the probability that each of the K clusters contains exactly one μ_k? Compute the numerical values of this probability for $K = 2, 5, 10$. Assume sampling is *with replacement* (a simplification that everyone uses in practice).

c. Where is $K \ln K$ coming from? CS 680 students only Now you pick $K' = cK \ln K$ centers at random i.e. $\mu_{1:K'} \sim uniform(\mathcal{D})$ (still with replacement), where $c > 1$ is a constant. Calculate the probability that the first cluster contains none of the K' centers (another simplification), using the approximation $(1 + \frac{1}{z})^z \approx e$ for $z > 0$ large. What is the value you obtain (should be a simple formula depending on K and c).

Problem 4 – Mixture likelihod and gradient descent

We will consider the finite Mixture of Gaussians model defined by

$$f(x) \;=\; \frac{1}{K} \sum_{k=1}^{K} \mathcal{N}(x; \mu_k, \sigma_k^2 I_d)\,, \tag{2}$$

where $x, \mu_{1:K} \in \mathbb{R}^d$, $\sigma_{1:K}^2 \in (0, \infty)$, I_d is the unit matrix of order d, $\mathcal{N}(x; \mu_k, \sigma_k^2 I_d) = f_k(x)$ is the normal distribution, and $K > 1$ an integer.

The gradient w.r.t. $\mu_{1:k}, \sigma_{1:K}^2$ of the likelihood of a single point x under model $f(x)$ are respectively:

$$\frac{\partial f}{\partial \mu_k} \;=\; \frac{1}{K\sigma_k^2}(x - \mu_k)\mathcal{N}(x; \mu_k, \sigma_k^2 I_d)\,, \tag{3}$$

$$\frac{\partial f}{\partial \sigma_k^2} \;=\; \left[\frac{||x - \mu_k||^2}{2K(\sigma_k^2)^2} - \frac{n}{2\sigma_k^2}\right] \mathcal{N}(x; \mu_k, \sigma_k^2 I_d)\,. \tag{4}$$

For the rest of the problem, assume that we observe a sample $\mathcal{D} = \{x^1, \ldots x^n\}$.

4.a Write the explicit formula for the E step of the EM algorithm for this model.

4.b Write the M step of the EM algorithm for fitting the model in (2) to the data $\mathcal{D}$ by specializing the formulas in the lecture.

4.c Write now the log-likelihood of the data and derive the expression of the gradient of the log-likelihood w.r.t μ_k. Your final expression should be simple and depend on some variables from **a, b**.

4.d Denote by $\hat{\mu}_{1:K}$ the values of the centers $\mu_{1:K}$ at a local maximum of the log-likelihood. Show that $\hat{\mu}_k$ can be written as a weighted mean of the data, i.e

$$\hat{\mu}_k = \sum_i u_{ki} x^i\,, \tag{5}$$

with $u_{ki} \geq 0$ for all k, i and $\sum_i u_{ki} = 1$. Which equation in Lecture VI have you recovered?