Unsupervised Learning — Practice Questions

CM52054: Foundational Machine Learning
Practice set with fully worked answers

Part A - Concept & Short-Answer

1) **Supervised vs. unsupervised learning (notation).**
Define supervised learning and unsupervised learning using dataset notation.

Answer: Supervised learning assumes labelled training pairs

$$D = \{(x_i, y_i)\}_{i=1}^{N},$$

and learns a mapping from inputs to outputs. Unsupervised learning assumes only unlabelled inputs

$$D = \{x_i\}_{i=1}^{N},$$

and aims to discover latent structure or regularities in the data.

2) **What is clustering?**
Give a concise definition of clustering in unsupervised learning.

Answer: Clustering partitions a set of unlabelled samples into groups (clusters) such that samples within a cluster are more similar to each other than to samples in other clusters, according to a chosen similarity or distance criterion.

3) **Motivations and applications.**
State three motivations or applications of clustering discussed in the lecture (any three).

Answer: Examples include: (i) data generalisation and representing missing data via cluster centres; (ii) data compression by replacing many points with a small set of representatives; (iii) privacy-preserving sharing by communicating cluster centres rather than raw individual-level records.

4) **Good clustering criteria.**
State the two criteria for "good clustering" used in the lecture.

Answer: A good clustering aims for (i) low intra-cluster distances (compact clusters) and (ii) high inter-cluster distances (well-separated clusters).

5) **Trade-off in clustering objectives.**
Briefly explain why "low intra-cluster distance" and "high inter-cluster distance" can be in tension.

Answer: Making clusters extremely compact can lead to over-segmentation (too many clusters), while enforcing extreme separation can split continuous or naturally connected structures; practical clustering balances these competing aims.

Part B - Algorithms & Comparisons

6) **k-means: algorithm steps.**
State the two alternating steps of k-means and write the centroid update equation.

Answer:

a) *Assignment:* assign each point x_i to the nearest centroid.

b) *Update:* recompute each centroid as the mean of its assigned points:

$$c_k \leftarrow \frac{1}{n_k} \sum_{x \in C_k} x,$$

where C_k is the set of points assigned to cluster k and $n_k = |C_k|$.

7) **k-means: stopping criterion.**
What stopping criterion is used for k-means in the lecture?

Answer: Stop when the assignments (and thus the centroids) no longer change, i.e., the algorithm has converged.

8) **k-means: non-determinism.**
Why is k-means considered non-deterministic?

Answer: Because the final clustering can depend on the initialisation of centroids; different initial centres can lead to different local optima.

9) **k-means: advantages and disadvantages.**
Give two advantages and two disadvantages of k-means.

Answer:

- Advantages: simple and scalable; converges in a finite number of iterations.
- Disadvantages: must choose K in advance; sensitive to outliers and can perform poorly when clusters have varying densities/sizes or in high-dimensional feature spaces where distances become less informative.

10) **k-means: per-iteration complexity.**
Given N points, M features, and K clusters, state the per-iteration time complexity for (a) assignment and (b) centroid updates.

Answer: (a) Assignment requires computing distances to K centroids: $O(KNM)$.
(b) Updating centroids requires aggregating across points/features: $O(NM)$.

11) **k-means: one-iteration worked example (1D).**
Let $X = \{0, 2, 8, 10\}$ and initialise centroids $c_1 = 0$, $c_2 = 10$ using Euclidean distance. Perform one assignment + update iteration and report the new centroids.

Answer: Assignments: $\{0, 2\} \to c_1$ and $\{8, 10\} \to c_2$. Updates:

$$c_1 \leftarrow \frac{0+2}{2} = 1, \qquad c_2 \leftarrow \frac{8+10}{2} = 9.$$

12) **k-means: high-dimensional features.**
Explain briefly why k-means can be problematic in high-dimensional feature spaces.

Answer: k-means relies on distance computations; in high dimensions, distance evaluation is expensive and may become less discriminative, which can degrade clustering quality and efficiency.

13) **Hierarchical clustering: bottom-up vs top-down.**
Distinguish agglomerative (bottom-up) and divisive (top-down) hierarchical clustering.

Answer: Agglomerative clustering starts with each point as its own cluster and iteratively merges clusters; divisive clustering starts with all points in one cluster and iteratively splits clusters.

14) **Single-linkage naming.**
Single-linkage clustering is also known as what?

Answer: Nearest neighbour clustering.

15) **Hierarchical clustering: speed vs structure.**
According to the lecture, which approach is typically faster, and which tends to reflect global structure better?

Answer: Agglomerative (bottom-up) is typically faster; divisive (top-down) tends to reflect global structure better.

Part C - More Challenging Questions

16) **k-means local optimum due to initialisation.**
Provide a concrete scenario showing that k-means can converge to a poor solution due to an unfortunate initialisation. Briefly explain the mechanism.

Answer: Consider two well-separated dense groups (true clusters) with roughly equal size. If both initial centroids are placed inside the same dense group, the first assignment can split that group into two clusters while assigning the other group to the nearer of those two centroids. Subsequent mean-updates may keep both centroids in or near the first group (depending on geometry), yielding a locally optimal but semantically poor partition that does not align with the two-group structure. This illustrates sensitivity to initial centroids and convergence to local minima.

17) **Hierarchical clustering: linkage choice and cluster "chaining".**
Single-linkage (nearest neighbour) can exhibit a "chaining" effect. Explain what this means and why it can be undesirable.

Answer: In single-linkage, the distance between two clusters is defined by the closest pair of points across clusters. As a result, clusters can be merged through a sequence of near-neighbour connections, producing long, thin clusters that "chain" together even when the overall groups are not compact. This can be undesirable because it may merge distinct groups via sparse bridges and yield clusters that violate the intended notion of low intra-cluster distance.