Unsupervised learning

1. Compare and contrast supervised learning, unsupervised learning and reinforcement learning

   Ans. **Supervised learning:** Labelled training data to generalize labels to new instances

   Supervised learning takes the form of function approximation where you are given a bunch of x, y pairs and your goal is to find function f that will map some new x to a proper y.

   y=f(x) →find f for new x to map to a proper y

   -Function approximation

   Unsupervised learning: Making sense of unlabeled data. You are given a bunch of x and y and your goal is to find some function that gives you a compact description of the set of x's that you have seen.

   -Data description

   Reinforcement learning: Given a string of pairs of data, we have to learn some function to generate f. We are given x, z. y=f(x)

2. What is a clustering problem?

   Ans. Clustering problem is taking a set of objects and putting them into groups.

   Given: set of objects X and Inter-object distances D(x,y)=D(y,x) and $x,y \in X$

   The output that a clustering algorithm needs to create is a partition.

   Output: partition $P_D(x)=P_D(y)$ if x and y is in same cluster

3. What does a clustering algorithm have in common with K-nearest neighbors algorithm?
   Ans. In k-NN, the domain knowledge is distances, similarity between objects. It is like clustering as they depend on similarities.

4. Difference between single linkage clustering, average linkage clustering and maximum linkage clustering? **Ans.** intercluster distance: single linkage: closest, average linkage: mean and maximum linkage: furthest.

   Ans. K means is similar to the single linkage but rather than using a neighbor to neighbor approach it uses a **means.**

5. How is median linkage clustering being different from average/mean linkage clustering?
   Ans. Median is a non-metric statistics, the ordering of the numbers matters whereas mean is a metric statistics, the details of the numbers matter a lot.
6. How/why single linkage clustering is a hierarchical agglomerative clustering structure?
   Ans. Hierarchical clustering seeks to build a hierarchy of clusters. Agglomerative or bottom up approach: initially each object is considered a cluster and two closest clusters are merged as one moves up the hierarchy.
7. How do we define domain knowledge in a clustering problem?

Ans. how we define intercluster distance is the domain knowledge in a clustering problem.

8. Given n as points/objects, k as clusters, what is the run time for single linkage clustering? **$O(n^3)$ explanation:** repeat k times (n/2), look at all distances to find the closest pair $O(n^2)$ with different labels.
9. True or false, single linkage clustering is deterministic. **True**. We run the same algorithm and unless there's ties in the distances, it gives us an answer.
10. True of false single linkage clustering requires randomized optimization. **False.** It does not require any randomized optimization.
11. True or false**,** k-means always converge and never stuck. **False. Can get stuck in local optima**
12. Given error defined as how far an object is from the center of its cluster, in metric distance, can k-means always improving and converge?

    Ans. Error cannot go up, it will always go down or remain the same otherwise Argmin won't select it. Partition of object x is Pt(x)=argmin(x-center of cluster). Monotonically non-increasing in error.

    There is a finite number of configurations in **K-means** and as you are not getting worse in error matrix, as long as you have some way of breaking ties, eventually it stops and gets convergence.
13. Given the monotonically non-decreasing likelihood property of EM, why it does not converge in theory?

    Ans. In **EM**, there **are infinite number of probabilities** (configurations) and you don't do worse but you keep getting closer every single time. Since there is an infinite number of configurations, the step by which you get better keep getting smaller. So you never really approach the final best configuration.
14. True or false, EM will not diverge. **True**. It will not diverge.
15. True or false EM will work with Gaussian distribution only. **False**. It is an algorithm that can work anytime we work with probability distribution. There are many different algorithms that work in different scenarios by defining different probability distributions and we just need to figure out the E step and M step.
16. True or false EM does not get stuck in local optima. **False**. It can get stuck in local optima with a randomized optimization and then we need to **randomly restart.**
17. True or false No clustering algorithm have all three desirable properties: richness, scale-invariance, and consistency. **True.** These properties are mutually contradictory in a sense as defined by impossibility theorem.

    Richness means any way of clustering could be an output. Scale-invariance: does not care about distance. Consistency: shrinking intra/expanding intercluster distance does not change clustering
18. Define a clustering algorithm that has the desirable properties richness and scale-invariance

    Ans. Assume single link clustering and we are going to keep doing clusters until we merge clusters that are more than theta/w units apart where theta=a parameter and w=largest pair-wise distance over the entire dataset. Here it is rich as we can move the points around and have any number of clusters. Then we have scale invariance as we can if the maximum distances are bigger, then w will be bigger by exactly the same amount so will end up back where we were. So distance will be same as before and so it's scale invariant. But if we chance w, then clustering will change.

19. Define a clustering algorithm that has the desirable properties richness and consistency

    Ans. Assume single link clustering and we are going to have some parameter theta and not stop until the clusters are theta units apart. For example: theta=10 ft. If all the points are 10 ft within one another, 1 cluster, outside 10 ft, n clusters. So we can have any number of clusters. so it is rich. Since it depends on distance, it is not scale-invariant. But it does not depend of the intra-cluster distance or inter-cluster distance as they can be theta ft apart or more than theta ft apart. So there is consistency.

20. Define a clustering algorithm that has the desirable properties scale-invariance and consistency

    Ans. We have n items and we are going to stop when we have n/2 clusters in single link clustering. Since it has a fixed number of clusters, it does not have a richness property because richness would allow us to have any number of clusters. But it does not care about the scale of distances as we will get the same clusters even if we multiply the distance by 2. So there is scale-invariance. And by same logic they are consistent.

21. True or false. k-means clustering algorithm has the desirable properties scale-invariance and richness. **True.** Consistency means introducing a new distance function would not change the cluster configuration. K-means does not have consistency as it is centroid based.
22. Compare and contrast k-means and EM clustering algorithms

    Ans. K-means: hard assignment, all variance same | EM: soft assignment, variance can be different

23. Explain how k-means is a special case of EM (PS2)

    EM algorithm finds a Gaussian mixture model to fit the data. K-means procedure is a special case of Gaussian mixture model where you don't consider the variance between the samples. In fact a variance of K-means, called the Lloyed algorithm, is essentially using EM algorithm with a centroid model.

24. Explain when EM would be a better choice for clustering than k-means?

    Ans. Only EM has the capability of handling overlapping clusters.