Supervised vs. Unsupervised learning

1. Discuss the differences between supervised and unsupervised learning.

   Solution. Answers will vary, some key points are:

   (a) Input data and labels: In supervised learning we have a set of features and a label we are trying to predict. In unsupervised learning we don't have a label we can use to refine the model.

   (b) Following from (1), since we don't have labeled training so there is no quality metric we can use to validate the model.

   (c) (Stretch) Randomness/initialization of the model has a large impact on the model's final state. This makes it difficult to systematically find "good" models.

2. For the machine learning algorithms we've discussed (linear regression, ridge regression, LASSO, logistic regression, decision trees, k-nn (document retrieval), k-means, hierarchical clustering), classify whether they are a supervised or unsupervised learning algorithm.

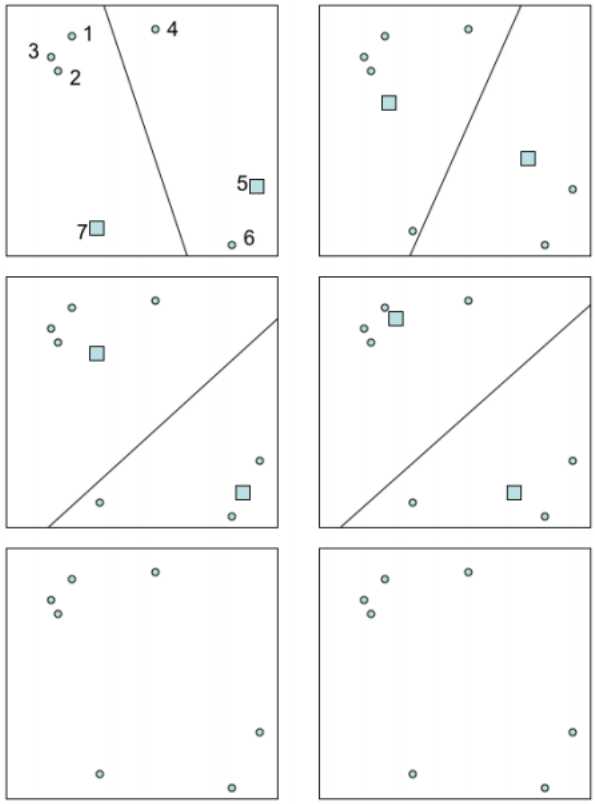
   Solution.
   Supervised Algorithms: Linear regression, ridge regression, LASSO regression, logistic regression, decision trees
   Unsupervised Algorithms: k-nearest neighbors, k-means, hierarchical clustering.

k-means clustering

1.
```
procedure k-means:
create k initial clusters

while the algorithm has not converged:
    assign each point to its nearest centroid
    update centroids to be the center of all points in cluster
end
```

2.



2. Compare the merits and drawbacks of k-means to hierarchical clustering with regards to the following:

   (a) Efficiency

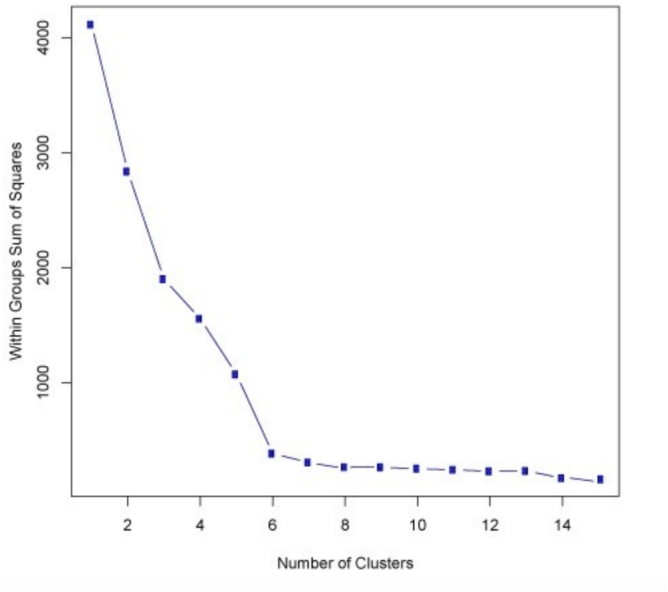
   Solution. k-means is more efficient in general than hierarchical clustering.

   (b) Hyper-parameters

   Solution. k-means requires picking k before the algorithm begins, whereas you can pick clusters for hierarchical after the algorithm has run. However, you must still pick a distance metric for hierarchical before starting.

3. Given the following graph, what is the common default for the number of clusters for our k-means algorithm?

(a)



Solution. 6 is the optimal number of clusters. Recall from lecture that cluster heterogeneity decreases monotonically as k approaches n. Therefore we want to pick a value of k such that heterogeneity is low but does not decrease by a trivial amount with more clusters.

(b) True or false: between two iterations of the k-means algorithm it is possible that no points are assigned to different clusters. Justify your answer.

Solution. True. Consider the state of cluster assignments once the algorithm has reached a local minima. The centroid will not move and all points will be classified the same between iterations.

Hierarchical Clustering

1. Suppose that the following distance matrix is given for 6 objects:

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

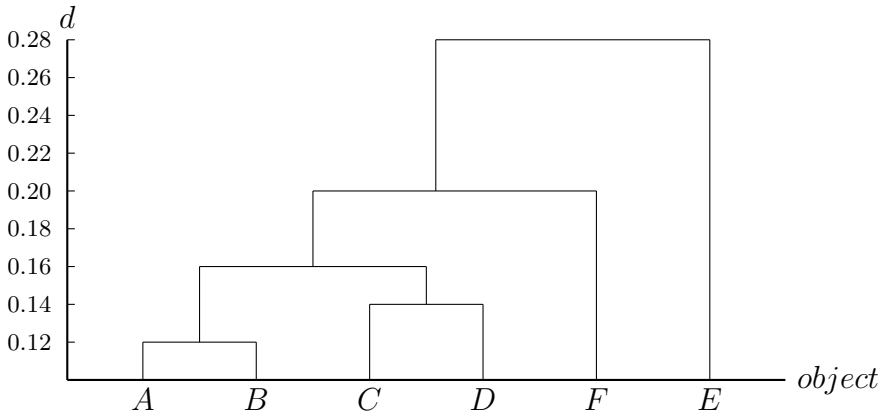
Single Linkage:

$$\min_{x_i \in C_1, x_j \in C_2} d(x_i, x_j)$$

Complete Linkage:

$$\max_{x_i \in C_1, x_j \in C_2} d(x_i, x_j)$$

(a) Show the final result of hierarchical clustering with single linkage by drawing a dendrogram.



(b) Show the final result of hierarchical clustering with complete linkage by drawing a dendrogram.