

## **Supervised vs. Unsupervised learning**

1. Discuss the differences between supervised and unsupervised learning.
  
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.

## k-means clustering

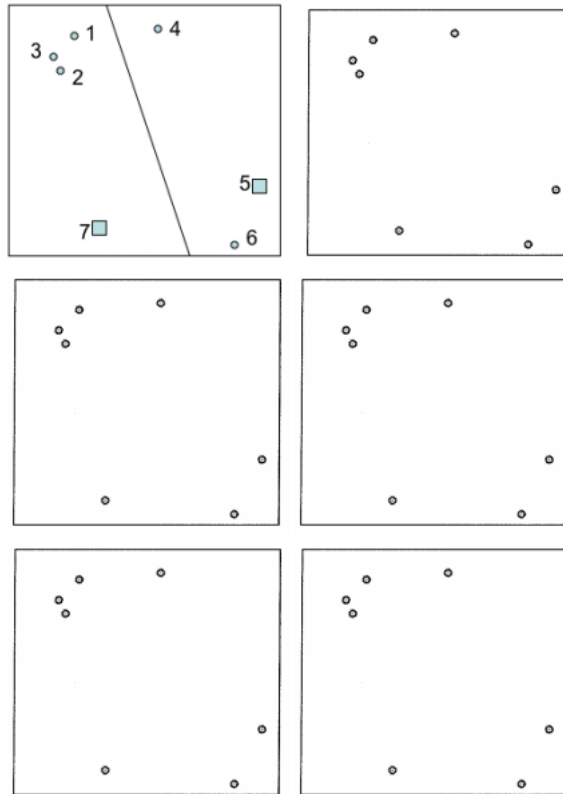
- Below we have provided partial pseudo-code for the k-means algorithm. Fill in the missing parts of the algorithm at locations marked (1) and (2).

```

procedure k-means:
  create k initial clusters

  while (1)
    assign each point to its nearest centroid
  (2)
end
  
```

- Perform k-means clustering on the dataset below. Assume that, initially, points 5 and 7 are chosen as centroids. Draw new centroids and decision boundaries until the algorithm converges.



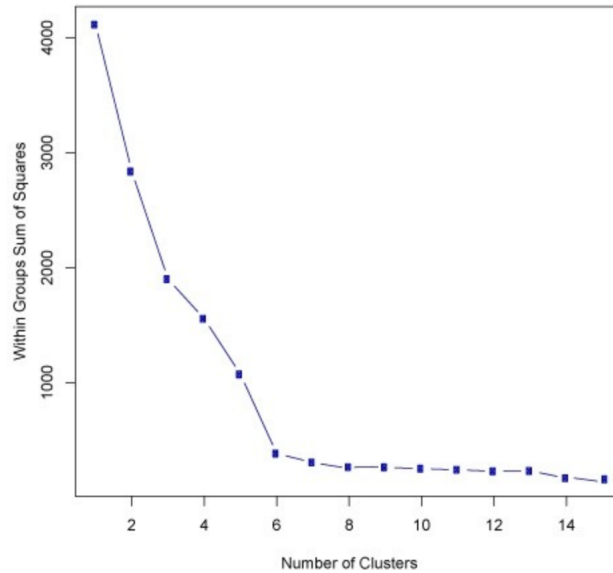
- Compare the merits and drawbacks of k-means to hierarchical clustering with regards to the following:

(a) Efficiency

(b) Hyper-parameters

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

(a)



(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.

# 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

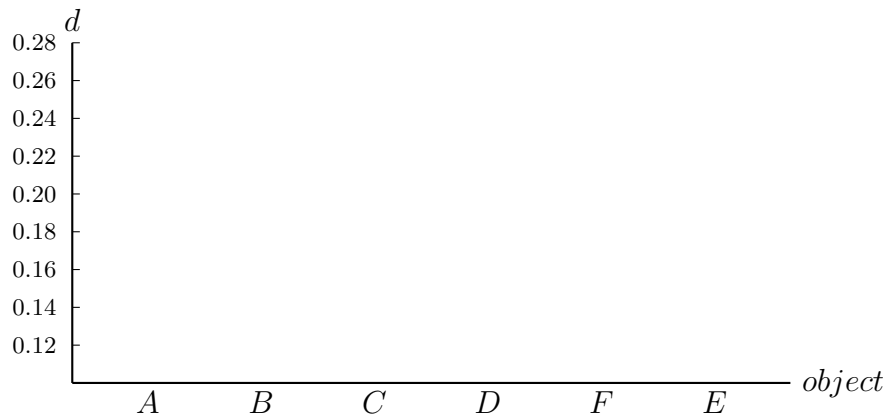
**Single Linkage:**

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

**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.

