
EECS 16ML Introduction to Machine Learning Skills

Fall 2020

Quiz

Test your Understanding of K-Means

1. What is the difference between supervised and unsupervised learning? Which category does K-Means fall under?
2. Cost is the counterpart to error for unsupervised, clustering models.
 - (a) Explain why the sum of squared distance is used to measure cost for K-Means.
 - (b) What are other potential measures of cost for K-Means?
3. Is K-Means always guaranteed to converge? Why or why not?
4. Will K-Means always produce the same final clusters each time it's run? Why or why not?
5. What is the big-O runtime of Lloyd's Algorithm? Let n = number of data points, d = dimension of each data point, k = number of clusters, and i = number of iterations needed till convergence.
6. Why are traditional hyperparameter tuning techniques, like grid search, ineffective for picking k ?
7. In Project Part 2, Example 1: What explains the phenomenon occurring with the clustering? What steps can you take to prevent this?
8. In Project Part 2, Example 3: What explains the performance of the clustering on the given dataset? What solves the issue and clusters the data as intended?
9. What are the benefits of Kernelized K-Means? Describe an appropriate situation where you would use Kernelized K-Means over regular K-Means.
10. How would you modify K-Means if you knew the measurement of certain features were more accurate than other features?
11. Another variation of K-Means is called K-Medoids. A medoid is defined as the object of a cluster whose average dissimilarity to all objects in the cluster is minimal - the most centrally located point in the cluster. In this variation, instead of updating the centroids to be the average of all the points in the cluster, the centroid is chosen to be the medoid of the cluster. What are the advantages and disadvantages of this method?