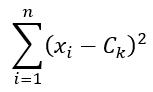
Chapter 1. Intro to Unsupervised learning

1. Which statement about unsupervised algorithms is TRUE?
   1. Unsupervised algorithms are relevant when we don’t have the outcomes we are trying to predict and when we want to break down our data set into smaller groups. Correct answer. Correct! They are helpful to find structures within our data set and when we want to partition our data set into smaller pieces for a better performance.
2. Which of these options is NOT an example of Unsupervised Learning?
   1. Explaining the relationship between an individual's income and the price they pay for a car. Correct answer Correct! This is an example best suited for regression, which is a supervised learning model.
3. What is one of the real-world solutions to fix the problems of the curse dimensionality?
   1. Reduce the dimension of the data set Correct! By doing dimensionality reduction we can improve both the performance and the interpretability of this grouping.
4. Which of the following examples is NOT a common use case of clustering in the real world?
   1. Determine risk factor and prevention factors for diseases such as osteoporosis. This is a Multiple logistic regression use case applied to clinical research.
5. Which statement is a common use of Dimension Reduction in the real world?
   1. Image tracking. This is an example of reduce data to the primary factors.

Quiz 2 K-means

1. ‘We initialize our k-means algorithm by taking 2 random points and these points are going to act as the centroids”
   1. True
2. Which of the following statements best describes the iterative part of the k-means algorithm
   1. The k-means algorithm adjusts the centroids to the new mean of each cluster, and then it keeps repeating this process until no example is assigned to another cluster
3. ‘The problem with k-means algorithm is that is sensitive to the choice of the initial points, so different initial configurations may yield different results.”
   1. True
4. Which statement describes better ‘the smarter initialization of k-means clusters’
   1. Pick one point random as initial point and for the second pick instead of doing ti randomly we prioritize by assigning the probability of the distance.
5. What happen with our second cluster centroid when we use the probability formula.
   1. When we use the probability formula, we put more weight on the points that are far away. So our second cluster centroid is likely going to be more distant.

Final

1. (True/False) K-means clustering algorithm relies on finding clusters centers to group data points based on minimizing the sum of square errors between each data point and its cluster centroid.
   1. True
2. What’s the name of the default initialization for K-means?
   1. K-Means++
3. What is the implication of a small standard deviation of the clusters?
   1. The standard deviation of the cluster defines how tightly around each one of the centroids are. With a small standard deviation, the points will be closer to the centroids.
4. After we plot our elbow and we find the inflection point, what does that point indicate to us?
   1. The ideal number of clusters.
5. (True/False) We can use K-means to reduce the size of high-quality images by just keeping the important information and grouping the colors with the right number of clusters.
   1. True
6. What is one of the most suitable ways to choose K when the number of clusters is unclear?
   1. By evaluating Clustering performance such as Inertia and Distortion.
7. Which statement best describes the formula for Inertia?
   1. The Sum of squares distance from each point (xi) to its clusters (ck)
   2. 
8. *Which statement describes correctly the use of distortion and inertia?*
   1. When the similarity of the points in the cluster are more important you should use distortion and if you are more concern that clusters have similar numbers of points then you should use inertia.
9. Select the approach that can help you find the cluster with best inertia
   1. Compute the resulting inertia or distortion, keep the results, and see which one of the different initializations of configurations lead to the best inertia or distortion. As an example of this, the best inertia result is the **lowest** value.
10. Which method is commonly used to select the right number of clusters?
    1. The elbow method.

Week 2

1. *(True/False) Is the following statement true or false? “Our choice of Distance Metric will be extremely important when discussing our clustering algorithms and to clustering success.”*
   1. True
2. *What is the other name we can give to the L2 distance?*
   1. Euclidean Distance
3. *Which of the following statements is a business case for the use of the Manhattan distance (L1)?*
   1. We use it in business cases where there is very high dimensionality.
4. *What is thekey feature for the Cosine Distance?*
   1. The Cosine Distance, which takes into acount the angle between 2 points.
5. *The following statement is an example of a business case where we can use the Cosine Distance?*
   1. Cosine is better for data such as text where location of occurrence is less important.
6. *Which distance metric is usefulwhen we have text documents and we want to group similar topics together?*
   1. Jaccard
7. (True/False) Hierarchical Agglomerative Clustering algorithm will try to continuously split out, and merge new clusters successively until it reaches a level of convergence.
   1. True
8. Why we need a stopping criterion when we are using the HAC?
   1. The algorithm will turn our data into just one cluster.
9. (True/False) Does the following statemen is a TRUE or FALSE explanation about the key operation of the DBSCAN algorithm? “A key part of this algorithm is that truly finds clusters of data rather that partitioning it, works better when we have noise in our data set and properly find the outliers…”
   1. True
10. According to the DBSCAN required inputs, which statement describes the n\_clu input?
    1. Determines density threshold (for fixed Ɛ) (The minimum amount of points for a particular point to be consider a core point of a cluster).
11. How do we define the core points when we use the DBSCAN algorithm?
    1. A point that has more than n\_clu neighbors in their Ɛ-neighborhood.
12. *Which of the following statements is a characteristic of the K-means algorithm?*
    1. To determine the number of clusters we use the elbow method.
13. *Which of the following statements is a characteristic of the DBSCAN algorithm?*
    1. Can handle tons of data and weird shapes.
14. *Which of the following statements is a characteristic of the Hierarchical Clustering (Ward) algorithm?*
    1. It offers a lot of distance metrics and linkage options.
15. *Which of the following statements is a characteristic of the Mean Shift algorithm?*
    1. Not require us to set the number of clusters, the number of clusters will be determined for us.
16. When using DBSCAN, how does the algorithm determine that a cluster is complete and is time to move to a different point of the data set and potentially start a new cluster?
    1. When no point is left unvisited by the chain reaction.
17. Which of the following statements correctly defines the strengths of the DBSCAN algorithm?
    1. No need to specify the number of clusters (cf. K-means), allows for noise, and can handle arbitrary-shaped clusters.
18. Which of the following statements correctly defines the weaknesses of the DBSCAN algorithm?
    1. It needs two parameters as input, finding appropriate values of Ɛ and n\_clu can be difficult, and it does not do well with clusters of different density.
19. (True/false) Using the Single Linkage method with HAC helps you ensure a clear separation between clusters.
    1. True
20. (True/false) Does complete linkage refers to the maximum pairwise distance between clusters?
    1. True
21. Which of the following measure methods computes the inertia and pick the pair that is going to ultimately minimize the inertia value?
    1. Ward linkage
22. Select the option that best completes the following sentence: For data with many features, principal components analysis
    1. generates new features that are linear combinations of the original features.
23. What is the main difference between kernel PCA and linear PCA?
    1. Kernel PCA tend to preserve the geometric distances between the points while reducing the dimensionality of the space.
24. (True/False) Multi-Dimensional Scaling (MDS) focuses on maintaining the geometric distances between points.
    1. False
25. (True/False) In some applications, NMF can make for more human interpretable latent features.
    1. True
26. Which of the following set of features is the least adapted to NMF?
    1. Monthly returns of a set of stock portfolios.
27. (True/False) The NMF can produce different outputs depending on its initialization.
    1. False
28. Which option is the dense representation of the matrix below?
    1. [[1 0 0 0],
    2. [0 3 0 0],
    3. [0 2 0 0],
    4. [0 0 4 2]]
29. When we use the DBSCAN algorithm, how do we know that our cluster is complete and is time to move to a different point of the data set and potentially start a new cluster?
    1. When no point is left unvisited by the chain reaction.
30. Which of the following statements correctly defines the strengths of the DBSCAN algorithm?
    1. No need to specify the number of clusters (cf. K-means), allows for noise, and can handle arbitrary-shaped clusters.
31. Which of the following statements correctly defines the weaknesses of the DBSCAN algorithm?
    1. It needs two parameters as input, finding appropriate values of Ɛ and n\_clu can be difficult, and it does not do well with clusters of different density.
32. (True/False) Using the Single Linkage method with HAC helps you ensure a clear separation between clusters.
    1. True
33. (True/false) Does complete linkage refer to the maximum pairwise distance between clusters?
    1. True
34. Which of the following measure methods computes the inertia and picks the pair that is going to ultimately minimize the inertia value?
    1. Ward linkage