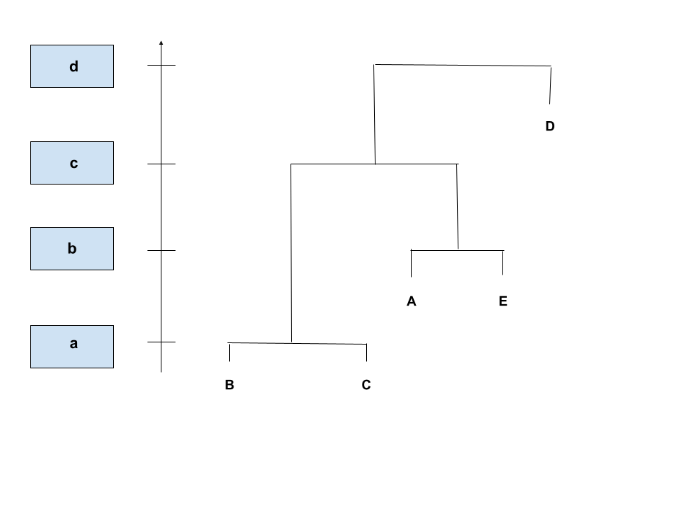
## Week 1 - PCA

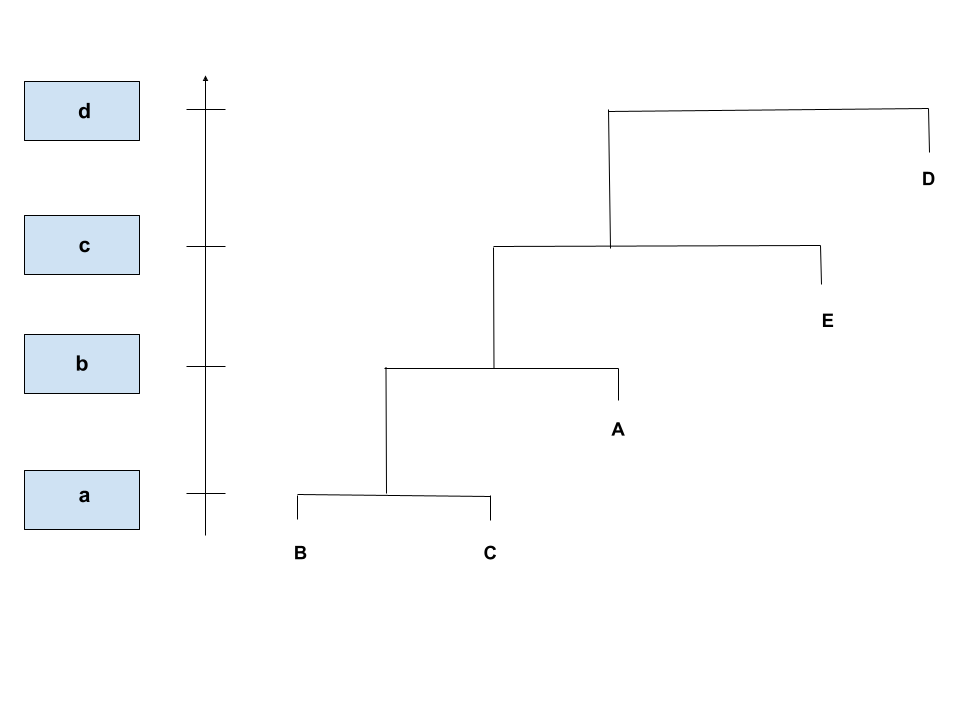
1. The main purpose of Unsupervised Learning is predicting either a continuous variable or a discrete variable.
   1. False. (Predicting either a continuous or discrete variable is the purpose of supervised learning. The primary purpose of unsupervised learning is to discover interesting hidden patterns in unlabeled data.)
2. The various phenomena that might arise during analysis of data in high dimensional space include the following:
   1. High dimensional data tend to become more redundant (correlated). (Consider the curse of dimensionality. In a sense, many dimensions will be redundant with high-dimensional data meaning that there are correlated attributes.)
   2. The data might become more and more sparse. (In a high-dimensional space, the data might become more and more sparse. For example, consider a matrix of user reviews for movies. If there are thousands of movies, each user has probably reviewed a small number of films.)
   3. The model trained using high dimensional is likely to overfit. (In a high-dimensional space, the data might become more and more sparse. With sparse data, there is an increased chance of overfitting.)
3. Which of the following can be a method to choose principal components in PCA?
   1. The components that preserve maximum variance.
   2. Choosing an axis that minimizes mean squared distance between the original dataset and its projection onto axis.
4. PCA is a linear feature extraction technique, performing a linear mapping of data to a lower-dimensional space.
   1. True.
5. Let the data be reduced to 50 dimensions using PCA (from dimensional space greater than 50). Which of the following statements are true?
   1. The variance captured by 16th principal component is greater compared to that of 19th principal component.
   2. The variance captured by 3rd principal component is greater compared to that of 47th principal component.
   3. The variance captured by 10th principal component is greater compared to that of 19th principal component.
6. With PCA, the eigenvalues obtained are real and non-negative.
7. True. (Eigenvalues represent the total variance that a given principal component can explain. Eigenvalues are real numbers. In theory, they can be positive or negative. In practice, though, they explain variance (every variance that isn't zero is positive.))

## Week 2 – Clustering

1. A dendrogram using single linkage is always shorter than a dendrogram using other types of linkage
   1. True. (The single linkage method uses the minimum value.)
2. k-means clustering leads to a unique solution.
   1. False. (k-means clustering finds local minima and its result may be different depending on the initialisation.)
3. Hierarchical clustering does not require any (hyper)parameters when finding the number of clusters.
   1. False. (Although hierarchical clustering doesn’t need to know the number of clusters in advance, it needs a threshold to cut the dendrogram and find the number of clusters.)
4. K-means clustering may also be used with cosine similarity metric.
   1. True. (Although Euclidean distance is the most popular choice with k-means, it can also work with the cosine similarity metric.)
5. In this question, we will examine how different linkage choices lead to hierarchical clustering dendrogram results. Consider a dissimilarity matrix as shown below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** |
| **A** | 0. | 0.74415444 | 0.43394511 | 0.88827404 | 0.5072831 |
| **B** | 0.74415444 | 0. | 0.33333333 | 0.60570207 | 0.75 |
| **C** | 0.43394511 | 0.33333333 | 0. | 0.906949 | 0.72918682 |
| **D** | 0.88827404 | 0.60570207 | 0.906949 | 0. | 0.89784658 |
| **E** | 0.5072831 | 0.75 | 0.72918682 | 0.89784658 | 0. |



1. What is the height value at the first linkage, represented by a if we use complete linkage?
   1. 0.333333. (Dissimilarity means the smaller number is similar pair. B-C pair is the most similar so it forms the first merge. It forms the first linkage at the height 0.333.)
2. 

What is the height value at the first linkage, represented by a if we use single linkage?

* 1. 0.33

## Week 3 – Recommender Systems

1. What would be the calculation complexity (cost) to compute the **Item-Item similarity matrix**?
   1. . (Your answer for the complexity is bounded by Number of items \* Number of items \* Number of users. This Big O notation complexity makes sense for a Item-Item similarity matrix.)
2. Recommendations or Predictions are made based on the properties (feaures) of an item in: Collaborative Filtering or Content Filtering.
   1. Content Filtering. (For Content Filtering, recommendations or predictions are made ***based on the properties (features) of an item***. This is different from Collaborative Filtering where even Item-basd collaborative filtering looks at items an user liked and then makes recommendations or predictions based on items similar to that item.\_
3. Consider a table with 4 user ratings on 8 items as shown below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **userID/game** | **WatchDog** | **GTAV** | **PUBG** | **StarCraft** | **CallOfDuty** | **FIFA** | **WOW** | **MineCraft** |
| firechicken | 3 | 3 | 4 |  | 4 | 2 | 3 |  |
| mike0702 | 3 | 5 | 4 | 3 | 3 |  |  | 4 |
| zephyros |  | 4 |  | 5 |  |  | 2 | 1 |
| dadvador | 2 |  |  | 4 |  | 4 | 4 | 5 |

Treating the utility matrix as boolean (buy or not buy), compute the Jaccard distance between firechicken and mike0702.

* 1. 0.5

1. Treating the utility matrix as boolean (buy or not buy), compute the Cosine distance between zephyros and dadvador.
   1. 0.671
2. Now, let's convert user ratings to good vs. bad. We consider it's good (1) if the rating is 3 or above, and bad (0) otherwise. Compute the Jaccard distance between firechicken and zephyros.
   1. 0.167