CSCE470 Written Assignment 3 (5 Points!)
Due by 11:59 pm on November 8th

1. Short answers (**1 Point!** 1/2 Point per question):

   a). Compare hierarchical clustering with flat clustering in terms of application scenaria and constraints.

   b). What are the main factors that affect clustering results? Please separately list factors that affect K-means results and factors that affect hierarchical agglomerative clustering results.

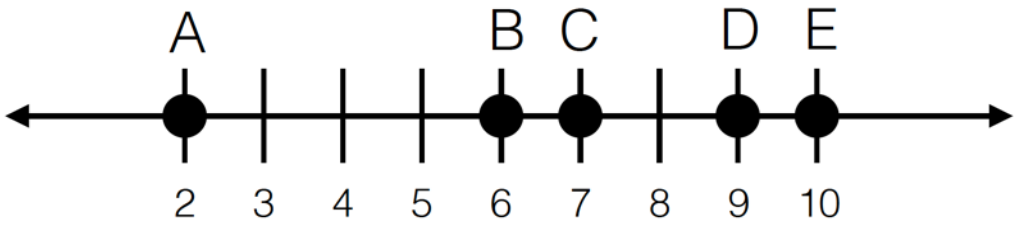
2. Problem Solving, text classification. (4 Points!).
   1. **K-means** (2 Point!)
      Suppose you have the following 6 documents:
      doc1: carp swim water
      doc2: horse neck land lung
      doc3: lion lung land
      doc4: elephant lung snout
      doc5: dolphins swim water lung
      doc6: seahorse swim water neck

      Represent each document as a raw term frequency vector, then normalize document vectors using Euclidean normalization. (**1 Point**!)

Assume k = 2 and the initial cluster centers are doc1 and doc2. Using Euclidean distance as your measure of distance, show us the first iteration of k-means. Make sure to clearly label your steps so that we can understand what you are doing. One iteration includes both 1) reassignment of documents (.**5 Point!)** and 2) recomputation of centroids **(.5 Point!).**

2. **Hierarchical Agglomerative Clustering** (2 Point!)

(a) Suppose you have the following five points (A, B, C, D, E) that you cluster using single-link hierarchical clustering. Draw the resulting dendrogram. (**1 Point!)**



(b) What if you cluster the five points using complete-link hierarchical clustering? Draw the resulting dendrogram. (**1 Point!)**