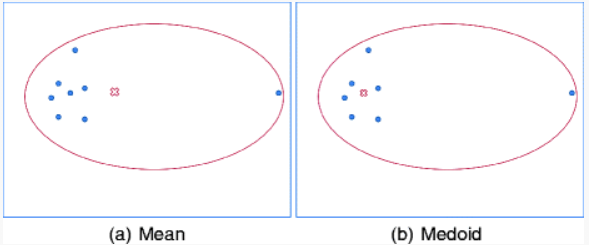
HW3, Machine Learning in Healthcare – 336546

By Nathan Berdugo

1. Clustering  
     
   a. Yes, K-medoid is indeed more robust to noise than the K-means algorithm. Whereas K-medoid and K-means algorithm share common characteristics (as they both try to minimize the distance between an example to a point in space), they have some intrinsic differences. Here I will explain the difference that makes K-means more sensitive to noise (and outliers) than K-medoid.  
     
   As explained in the question introduction, **K-medoid** algorithm uses L1 minimization in order to achieve dissimilarity minimization between the examples within the dataset. An important difference between K-means and K-medoid is that *in K-medoid, the center of a cluster is chosen from the actual dataset*, while K-means not necessarily picks a 'centroid' from within the dataset.  
     
   By determining the center of a cluster with a point in space which is not included in the dataset, **K-means** allows noise or extreme data to influence center-point selection, since the algorithm - as its name implies - is based on the mean. As we learned in the course lectures, as well as in other courses (especially in the Statistics 094423), **the mean is very sensitive to extreme values**. By using K-means, we give great weight to those noises and extreme values (outliers) that can significantly affect the determination of the cluster center.

This effect will be reflected to a lesser extent in the K-medoid algorithm since as stated, the center is determined from the dataset and therefore, noise nor outliers will not be reflected in this method.

**Figure 1** – A demonstration of a 2D-dataset with an extreme value (on the right edge of the ellipse). One can notice that in K-means (a), the centroid is highly influenced by the outliers as an intrinsic characteristic of the mean. While K-medoid (b) chooses an example **from within the dataset** as the center of the cluster. 1

1/ Jin X., Han J. (2011) K-Medoids Clustering. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston  
  
b. To-be-proved: For 1D case of K-means, the centroid  which minimizes the term  is the mean of examples.  
  
  
  
Proof:  
  


One can notice that in order to minimize the term above, should be valued as  (i.e. ).

**Bonus.** To-be-proved: For K-medoid, the medoid which minimizes the term  is the median of examples, given that belongs to the dataset.

Proof (in next page)2:

# 2/ [The median minimizes the sum of absolute deviations (the ℓ1 norm)](https://math.stackexchange.com/questions/113270/the-median-minimizes-the-sum-of-absolute-deviations-the-ell-1-norm), StackExchange: https://math.stackexchange.com/questions/113270/the-median-minimizes-the-sum-of-absolute-deviations-the-ell-1-norm



1. SVM