Q1 – Clustering

a. K – means and K – medoid are similar clustering algorithms, except K – means relies on the mean of a cluster and K – medoid relies on the medoid sample of one. A medoid of a cluster is defined as the sample for which average dissimilarity between it and all the other samples of the cluster is minimal, somehow like the median. The major difference between the median of a group and the mean of a one, is that the median is not affected by outliers and random noise. For instance, considering the following array: [1,2,4,5,100].
Its median is 4 and its mean is 32, so obviously the median represents it better.
Indeed, the K – medoid algorithm is more robust since it is based on a value that resembles to the median much more than to the mean.

b. We wish to minimize the following:

$$f(\mu) = \sum_{i=1}^{m} (x_i - \mu)^2$$

Next, a derivation is needed:

$$\frac{d}{d\mu} f(\mu) = \sum_{i=1}^{m} 2(x_i - \mu) \cdot (-1) = (-2) \cdot \sum_{i=1}^{m} (x_i - \mu)$$

$$0 = (-2) \cdot \sum_{i=1}^{m} (x_i - \mu) \to \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} \mu$$

$$\sum_{i=1}^{m} x_i = m \cdot \mu$$

$$\mu = \frac{\sum_{i=1}^{m} x_i}{m} = mean(x_i)$$

Another derivation is needed:

$$\frac{d^2}{d\mu^2}f(\mu) = \frac{d}{d\mu}(-2)\cdot\sum_{i=1}^{m}(x_i-\mu) = \frac{d}{d\mu}\left((-2)\cdot\sum_{i=1}^{m}x_i + 2\cdot\sum_{i=1}^{m}\mu\right)$$

$$= \frac{d}{d\mu}(C+2m\mu) = 2m > 0$$

thus, it is a convex function and the global minima is the mean of x_i.

Bonus:

The main assumption is that m is an odd number so the median will be one of the samples. Verbal explanation for the even case is in the end of the question.

Now we wish to minimize the following function:

$$f(\mu) = \sum_{i=1}^{m}|x_i-\mu|$$

First, let us assume that there are a samples bigger than μ and b samples smaller than μ.

$$f(\mu) = \sum_{i=1}^{a}(x_i-\mu) + \sum_{j=1}^{b}(\mu-x_j) +^{**}(\mu-\mu)\;;[a+b=m-1]$$

$$** \;-\; \mu\ is\ one\ of\ the\ samples$$

If we show that when $a=b$ the function is minimized, then it means that the samples' number which is smaller than μ is equal to the samples' number that is bigger than μ and thus μ is the median.

$$f(\mu) = \sum_{i=1}^{a}x_i - \mu a + \mu b - \sum_{j=1}^{b}x_j$$

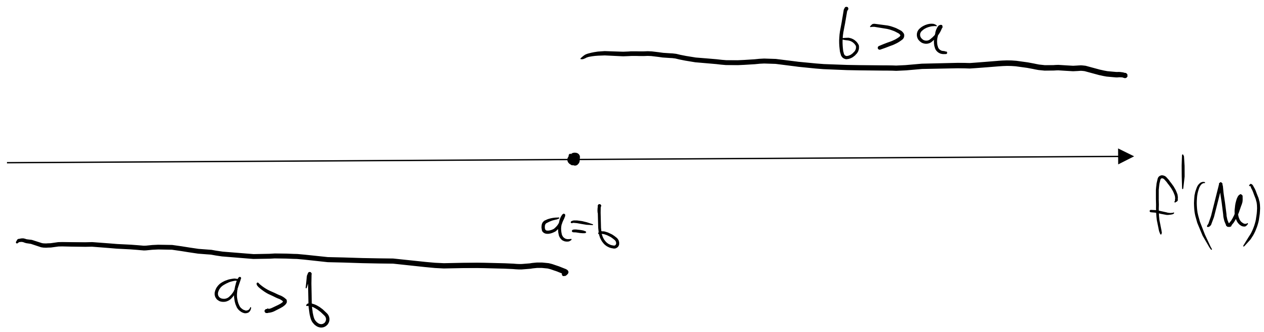
Derivation:

$$\frac{d}{d\mu} f(\mu) = -a + b$$

Compared to 0:

$$0 = -a + b \rightarrow a = b$$

We want to show that this is the global minima, for that we must understand how the derivative is dependent on a and b.



$$f'(\mu) = \begin{cases} C & b > a \\ 0 & a = b \\ -C & a > b \end{cases}$$

In conclusion, for $a = b$ we get global **minima** for $f(\mu)$, meaning that μ is indeed the median.

If m were an even number, we would choose the median as follows:

First, we divide m to two equal groups, arranged by their values.

$(\{x_{(1)}, x_{(2)}, \dots \dots, x_{(l)}, x_{(l+1)}, \dots \dots, x_{(m)}\})$

Then, the median is the mean of $x_{(l)}, x_{(l+1)}$ and the same calculation applies.