HW3

1.

a. The **K-means** clustering algorithm is **sensitive to outliers**, because a **mean** is easily influenced by **extreme values**. ... **Mean** is greatly influenced by the **outlier** and thus cannot represent the correct cluster center, while medoid is robust to the **outlier** and correctly represents the cluster center

b. Will define: $J = \sum_{i=1}^{m}(x_i - \mu)^2$

Will differentiate with respect to μ

And get: $\frac{\partial J}{\partial \mu} = \sum_{i=1}^{m} -2(x_i - \mu)$

Will substitute $\mu = \frac{\sum_i^m x_i}{m}$ and will get $= \sum_{i=1}^{m} -2\left(x_i - \frac{\sum_i^m x_i}{m}\right) =$

$$\left(-2(mx_1 - x_1 - x_2 - \cdots x_m) + \cdots + -2(mx_m - x_1 - x_2 - \cdots x_m)\right) will\ devide\ by\ -2\ and\ than\ will\ get = (m(x_1 + \cdots + x_m) - mx_1 - \cdots - mx_m) = 0$$

Therefor it's a extremum point will show it's a minimum rather than a maximum

Will differentiate again: $\frac{\partial^2 J}{\partial \mu^2} = 2m > 0$ so the 2nd derivative is always positive and thus the function is convex and have only minimum point.

c. Will define $J_2 = \sum_{i=1}^{m}|(x_i - \mu)|$ $f(x) = |x| \rightarrow f'(x) = \frac{x}{|x|}$ therefor

Will differentiate with respect to μ

And get: $\frac{\partial J}{\partial \mu} = -\sum_{i=1}^{m}\frac{(x_i - \mu)}{|(x_i - \mu)|}$

Will substitute $\mu = median(x)$ and will get $\frac{\partial J}{\partial \mu} = -\sum_{i=1}^{m}\frac{(x_i - median(x))}{|(x_i - median(x))|} =$

Will open: $-\frac{(x_1 - median(x))}{|(x_1 - median(x))|} - \cdots - \frac{(x_m - median(x))}{|(x_m - median(x))|} =$

$$-1\left(\frac{(x_1 - median(x))}{|(x_1 - median(x))|} + \cdots + \frac{(x_m - median(x))}{|(x_m - median(x))|}\right)$$

We know that exactly half of the x's bigger than the median and exactly half of them smaller therefore will get exactly half of the numerator are +1 and exactly half are -1 And all of the denominator are +1 there for will get something like that:

$$-1\big(1 + 1 + \cdots + (-1) + (-1)\big) = -1(0) = 0$$

Will differentiate again: $\frac{\partial^2 J}{\partial \mu^2} = 0$ so the 2nd derivative is always positive(or 0) and thus the function is convex and have only minimum point.

2. We have 2 SVM with linear karnel A and D . A as samples within the margins and thus have more soft margins and thus smaller penalty term C.

A-1

D-2

C and F are polynomial karnel, and its clear F is with higher degree.

F-4
C-3

B and E are both rbf, and B boundary is tighter so its with bigger γ,which is correlated with tighter boundary.
B-6
E-5

3.

a. The scientific term is generalization
b. 2p could imply about overfitting (lack of generalization) model if its 2 big.
   And 2ln(L) could imply about underfitting (lack of generalization) if its to small. Therefor a good AIC is if there is a balance between the 2 parameters.
c. If the balance is violated there could be under/overfitting.
d. AIC is a criterion that help you asses goodness of a model. Lower values of the index indicate the preferred model, that is, the one with the fewest parameters that still provides an adequate fit to the data.