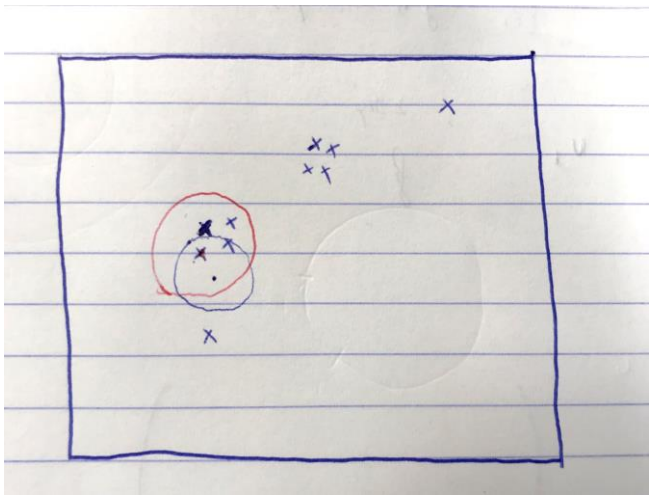
HW3 Jeremie Sasson 336319470

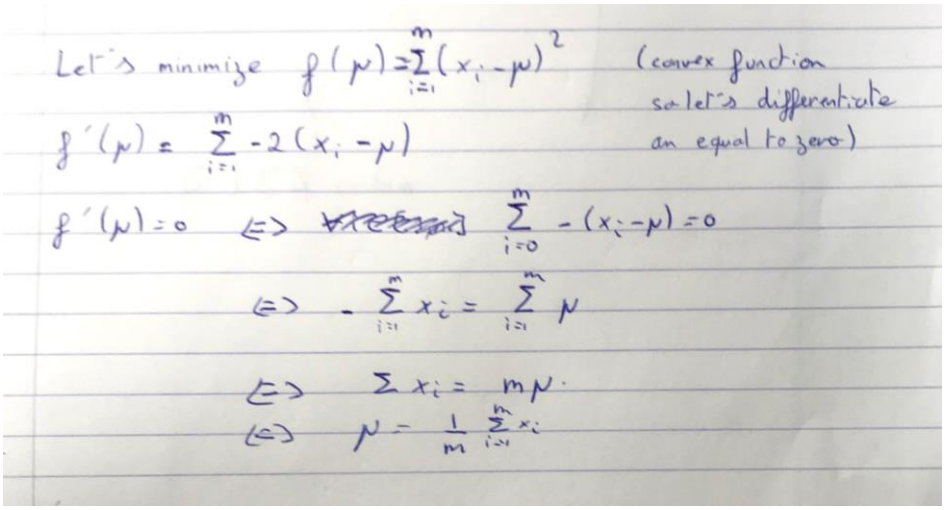
## 1) Clustering

a)
The k- medoid objective function seems to be more robust to noise. The fact that k-means enable centroid not to be part of the data make it more sensitive to noise than k-medoid. Let's take an exemple

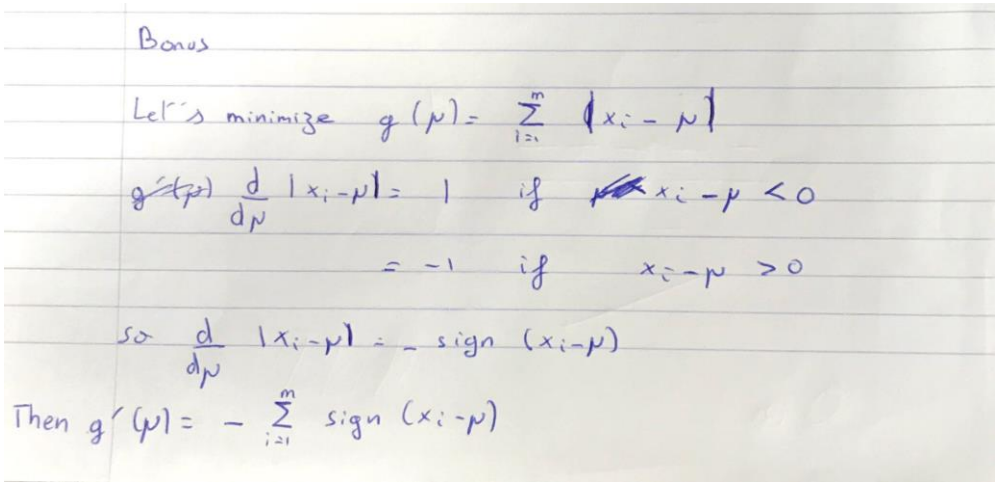


The blue circle will probably be the cluster result of a Kmeans algorithm and effectively it may be falsely biased by the noisy point.
Whereas the red circle may be the result of the k-medoid algorithm – biased by the noise but the bias is much less misleading.

b)




Bonus




The value of mu that zeroes this sum is such that half of xi are bigger than mu end the other half is smaller than mu.
Given that mu belong to the dataset, mu is actually, by definition, the median of the m examples.

## 2) SVM

Clearly A and D corresponds to Linear Kernel because of the shape of the threshold.
The more C is Big the more we penalize for mistakes small distance between threshold and points.
For sure the threshold in D tends to separates better the data in terms of margin so the algo penalizes more there

D: linear with C=1
A:linear with C=0.01
C and F corresponds to polynomial Kernel. C is more simple and its shape looks like a 2nd order polynomial. F is much more complex and clearly overfits – it is the 10 order polymomial.

C : 2nd order polynomial
F: 10th order polynomial

Gamma is a parameter that indicates how much we want to fit the training data. The higher Gamma the more we fit the training data. So too high of a Gamma can cause overfitting (from lecture). B tends to overfit, E appears to be more realistic.

B: RBF with gamma 1
E: RBF with gamma 0.2

## 3) Capability of generalization.

This Balance that Einstein speaks about is, in terms of machine learning the Bias Complexity trade off. "As simple as possible" → prevent overfitting. "But no simpler" → prevent underfitting.

2p is the number of parameters then increasing it leads to make out model more complexes. The Estimated likelihood that is log inversed make our criteria valorizes the algorithms as they are accurate. For a fixed p, The more our likelihood is big the more our AIC is small.
Thus, AIC tends to emphasizes accurate models and make our models more complex, but this complexity is over balanced by p (that appears in + sign).

As we said previously
Underfitting : our model is too simple -
Overfitting: out model is too complex -
Both prevent generalization.

As explained in b, AIC should be low.