Machine Learning in Healthcare
HW3

Aurelia Sebbane - 336334479

Question 1 - Clustering

a. Is K-medoid more robust to noise (or outliers) than the K-means algorithm? Explain your answer.

Yes, K-medoid is more robust to noise or outliers than the K-means algorithm. Indeed, K-means algorithm minimizes total square error in the Euclidian metric, therefore extreme values will affect the center position of the cluster in adequation with the mean, while K-medoid minimizes the distance between pair of points in the cluster.

b. Prove that for the 1D case ($x \in R^1$) of K-means, the centroid μ which minimizes the term $\sum_{i=1}^{m}(x_i - \mu)^2$ is the mean of m examples.

We are looking for minimal value, that means we need to derivate the term $\sum_{i=1}^{m}(x_i - \mu)^2$ by μ and to define it 0:

$$\frac{\partial}{\partial\mu}\left(\sum_{i=1}^{m}(x_i - \mu)^2\right) = 0$$

$$\rightarrow \cancel{-2m}\left(\sum_{i=1}^{m}(x_i - \mu)\right) = 0$$

$$\rightarrow \sum_{i=1}^{m} x_i - m\mu_{min} = 0 \rightarrow \mu_{min} = \frac{\sum_{i=1}^{m} x_i}{m}$$

We indeed found that the centroid which minimizes the term $\sum_{i=1}^{m}(x_i - \mu)^2$ is the mean of m examples.

Question 2 - SVM

- A and D match to a linear kernel SVM since they got a linear line to do the classification.
  The higher is C, the lower will be the margins. A has a lower margin than D so :
  <u>A → 2)</u>
  <u>D → 1)</u>

- The classification line that looks like most a 2nd order polynomial is the graph C (we can see a parabola) so:
  <u>C → 3)</u>

- RBF kernel fits to the graphs B and E since in RBF kernel we receive separation with a radius shape (polar nucleus, RBF = Radius Basis Function). The lower γ is, the higher the sample's influence is, which means :
  B → 6)
  E → 5)

- The shape of F looks unknown, we assume that it corresponds to a 10^{th} order polynomial kernel :
  F → 4)

Question 3 - Capability of generalization

a. What is the scientific term of the balance that Einstein meant to in machine learning aspect ?

The scientific term of the balance that Einstein meant to in machine learning aspect is: generalization = the balance between model performance and model complexity.

b. How does each of the terms ($2p, 2ln(\hat{L})$) in AIC affect the terms of the balance you defined in (a) ?

The higher is 2p, the higher the complexity will be, with p being the total number of learned parameters.
The higher is $\hat{L}$, the higher will be $2ln(\hat{L})$ and therefore the higher will be the model performance.

c. What are the two options that are likely to happen if this balance was violated ?

The two options that are likely to happen if this balance was violated are overfitting and underfitting.
Overfitting is likely to happen in the case of high performance and high complexity.
Underfitting is likely to happen in the case of low complexity.

d. What are we aiming for with the AIC ? Should it be high or low ? Explain

We would like to have the most simple and performant model. As we saw in (b), the lower is p, the lower is the complexity, and the higher is the likelihood, the higher are the model performances, which is exactly what we are lokking for. Wo we would like AIC to be low.