HW3 – BM-336546

Asaf licht

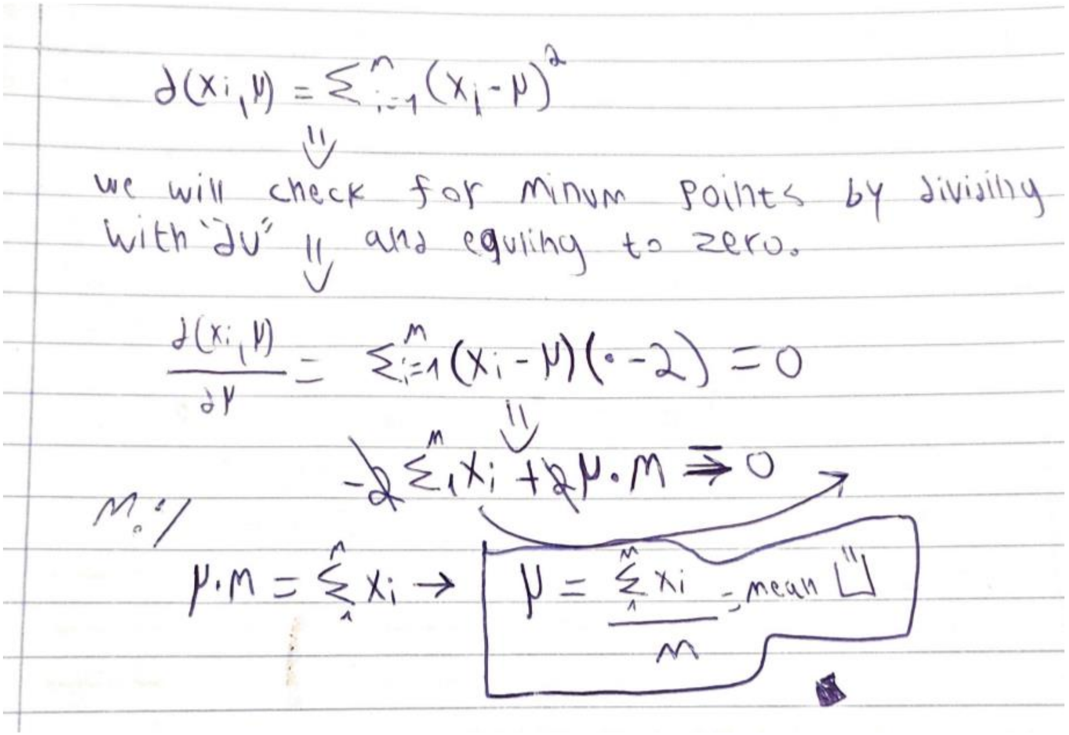
311129522

Q1

A . . "K-medoid" is more robust to noise and outliers then "K-means".

In K-means the centroid which we try to center the data according to,is given. The algorithem Assign each observation to the cluster with the closest mean calculated by the mean of data points selected randomly – and by measuring their squared distance from the centroid we apply our model. This model try to minimize the Squared distance due it is very sensitive to outliers and can compromised our model if it is not taken in consider. The ***k*-medoids algorithm** related to the **k-means** algorithm. Both the *k*-means and *k*-medoids algorithms are breaking the dataset up into groups. *K*-means attempts to minimize the total squred eror, while *k*-medoids minimizes the sum of dissimilarities between points labeled to be in a the same cluster and a point designated as the center of that cluster. In K-medoids the data point themselves is chosen to be the center.

B.



$$d(x_i,\mu) = \sum_{i=1}^{n}(x_i-\mu)^2$$

$\Downarrow$

we will check for minum points by dividing with "∂u" and equling to zero.

$\Downarrow$

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

$\Downarrow$

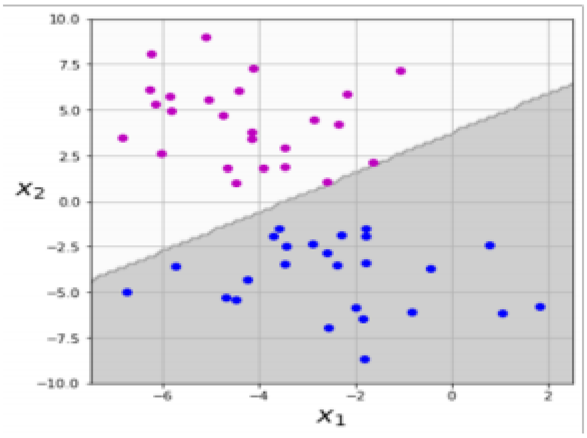
$$-\cancel{2}\sum_{1}^{m}x_i + \cancel{2}\mu\cdot m \Rightarrow 0$$

M:/

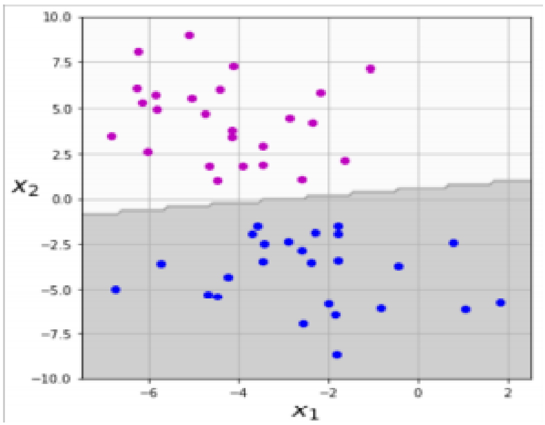
$$\mu\cdot m = \sum_{1}^{n}x_i \rightarrow \boxed{\mu = \frac{\sum_{1}^{m}x_i}{m} = \text{mean}}$$

Q2.

A



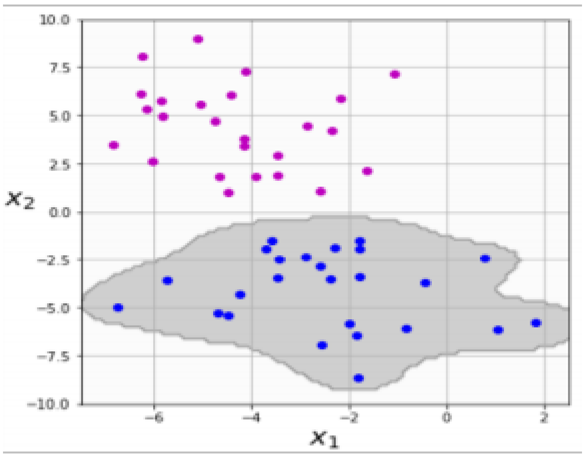
D



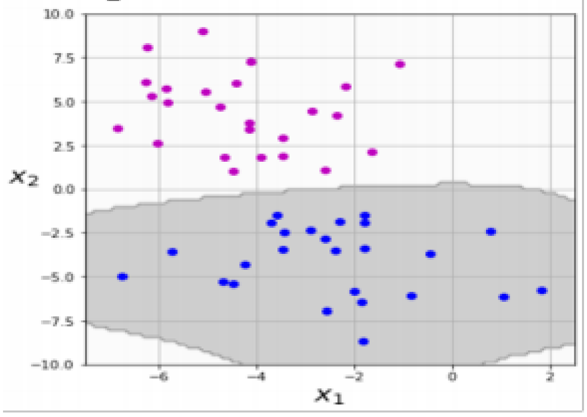
As we can see pretty clearly those 2 graphs represent a SVM with linear karnerl – as the separation between the groups are divided by a linear line, which we classifies our data according to.

the higher the c we choose the larger distance we will get between the extremists points of data to the linear line which dividing the data. So "A; kernel C= 0.01, and D ;kernel = 1) specific in A – the 2 points found on the line itself is probably in chance of mis-classification.

B

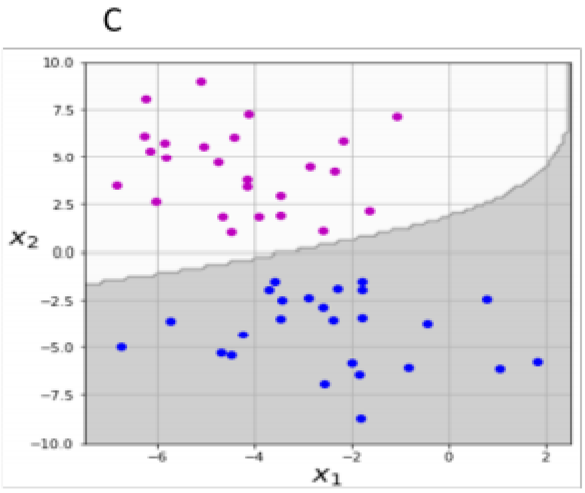


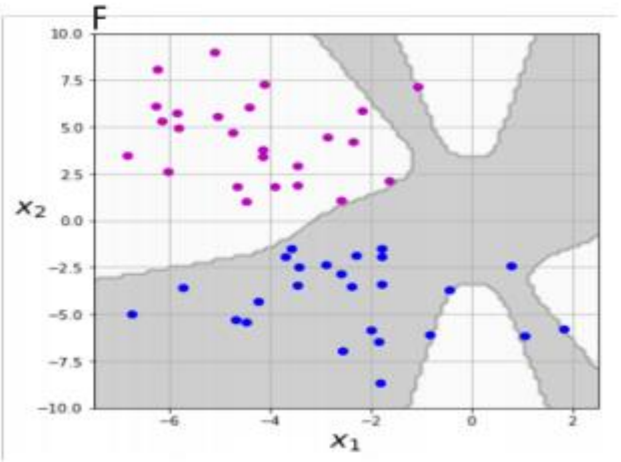
E



"A radial basis function (RBF) is a function whose value depends only on the distance between the input and some fixed points called a center".

$\gamma = \frac{1}{2\sigma^2}$. – so the higher gamma is the smaller the "spread" of the classifying function.
as we can see B is more "fitted" then E … so B; $\gamma = 1$ and E; $\gamma = 0.2$





We can see F graph is probably "overfitted" and the C graph is polynomial kernel from low degree. both are polynomial classifiers – but C is with second degree polynomial kernel and F is with 10 degree.

3.

A. the scientific term in machine learning is generalization. google defined : "generalization refers to your model's ability to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model." In this manner – if we will take a model to much complex– I will overfit and will not be able to preform properly when a new data is loaded.
**we will want to choose the parameters with logic thinking behind them – thse will give us a good fit but will be general enough to apply on new data.**

B. The parameter P is a measure for the complexity of the model – because it is the number of features, while L is the likelihood function that present the fit of the model. By using those 2 parameters we can balance the model – if we add a feature but the likelihood function value will not grow so we would prefer to keep the parameter outside of our model.

C. the balance is to find a good place between under fitting ( a low likelihood function) and over fitting (caused by an high likelihood value but complex model with not general enough features).

D. we are aiming to choose a minimal AIC. the + sign is referring to complexity so we will want it to be low. The – sign is referring to the likelihood function so we will want it to be high (for a good fit). A good model is one with high likelihood – coming out of low number of general features.