EXERCISE SESSION kernel K-Means: ADVANCED MACHINE LEARNING COURSE – EPFL – Lecturer A. Billard

a) Draw the partitioning of the space when using two datapoints in 2-dimension with an rbf kernel.
b) Do (a) with a polynomial kernel. Is the result affected by the placement of the datapoints?

Solution:

a) The question is to know whether there will be two clusters, each with one point, or one cluster with the two points and the other cluster empty:

- If we assume that the points are initialized each in one of the clusters ($x^i \in C^i, \forall i = 1,2$), we use the following equation to determine if a point (x^1) should join the other cluster (C^2) and form a two-point cluster:

$$\min_k d\left(x^i, C^k\right) = \min_k \left(\mathrm{k}\left(x^i, x^i\right) - \frac{2\sum_{x^j \in C^k} \mathrm{k}\left(x^i, x^j\right)}{m_k} + \frac{\sum_{x^j, x^l \in C^k} \mathrm{k}\left(x^j, x^l\right)}{\left(m_k\right)^2} \right)$$

With an RBF kernel, $\mathrm{k}\left(x^i, x^i\right) = 1$ and let's define $\mathrm{k}\left(x^1, x^2\right) = \mathrm{k}\left(x^2, x^1\right) = c$ with $c \in [0,1]$

$d\left(x^1, C^2\right) = 1 - \frac{2c}{1} + \frac{1}{1^2} = 2(1-c) \geq 0.$

Now let's compare the distance to the cluster C^1 of which x^1 is already part of:

$d\left(x^1, C^1\right) = 1 - \frac{2 \cdot 1}{1} + \frac{1}{1^2} = 1 - 2 + 1 = 0$

Therefore, $d\left(x^1, C^1\right) \leq d\left(x^1, C^2\right)$. We have equality when $\mathrm{c} = 1$ which is the case when $x^1 = x^2$.

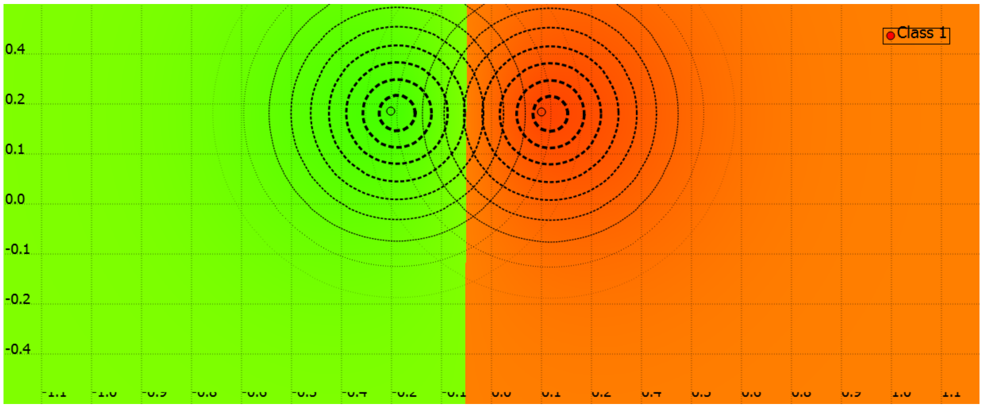
The solution is thus to always have one cluster for each point.
The partitioning of the space is symmetric between the two points (the median of the two points delimits the clusters).

- If the two points are initialized in the same cluster($x^i \in C^1, \forall i = 1,2$):

$d\left(x^1, C^2\right) = 1 - 0 = 1$

$d\left(x^1, C^1\right) = 1 - 2\frac{1+c}{2} + \frac{1+1+c+c}{4} = 1 - 1 - c + \frac{1+c}{2} = \frac{1}{2} - \frac{c}{2}$

The points will stay in the same cluster C^1 and the second cluster will be empty.
The isolines and the boundary look as follows:



b) For polynomial kernel, we have in the first case (each point assigned to each cluster):

$$d\left(x^1,C^1\right)=\left\|x^1\right\|^2-\frac{2\cdot\left\|x^1\right\|^2}{1}+\frac{\left\|x^1\right\|^2}{1^2}=0$$

$$d\left(x^1,C^2\right)=\left\|x^1\right\|^2-\frac{2\cdot\left\|x^1\right\|\left\|x^2\right\|\cos\left(x^1,x^2\right)}{1}+\frac{\left\|x^2\right\|^2}{1^2}\geq 0$$

It is zero when the two datapoints are identical.
The placement of the point in space plays no role.
The separation for homogeneous polynomial kernel of p=2 looks as follows: