Indian Institute of Technology, Madras
CS5691: Pattern recognition and Machine Learning
PRML Assignment-I Report

Ashutosh Kakadiya - CS18S013,
Rajan Kumar Soni - CS18S038

September 2019

Contents

Chapter 1

General Questions

1. Plot this data-set. What distribution might have generated this data-set? Why did you conclude so?
   - Scatter plot of Dataset-1 :

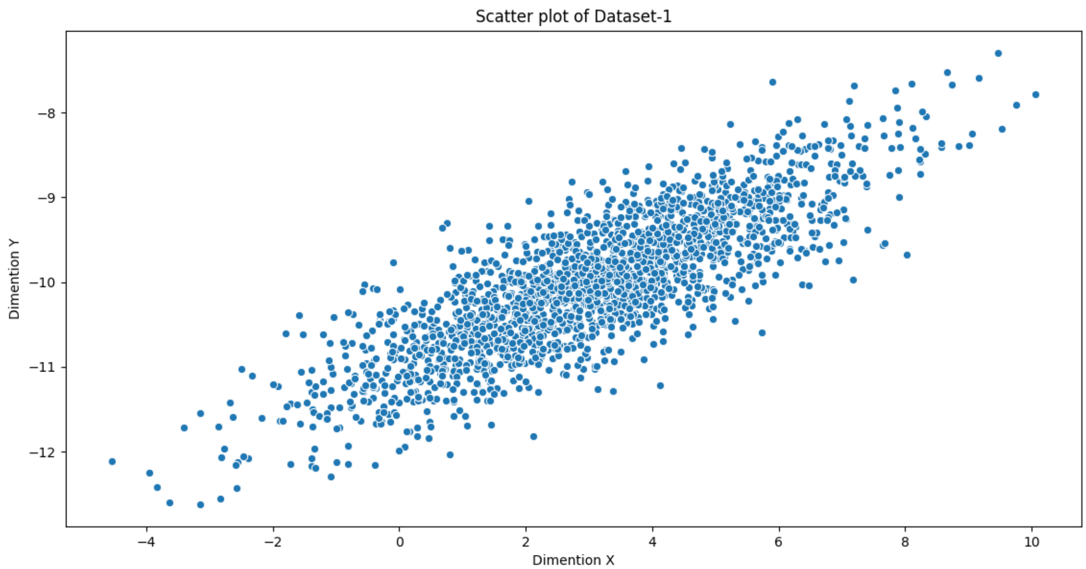


Figure 1.1: Scatter plot of Dataset-1

   We can clearly see the **elliptical contours of Bivariate Gaussian distribution of positive covariance matrix(General form matrix)**. Therefore **Bivariate Gaussian distribution** might generated this dataset.

2. Write a piece of code to obtain the maximum likelihood estimate of the parameters of the distribution that you think generated this data-set.

3. What is the log-likelihood value of observing this data-set for the parameters that you estimated in (2)?
   Ans. Log likelihood is : -5675.7157

4. Under Given assumption, What is the maximum likelihood estimator for μ?.
   Ans. μ value is not depend on co-variance matrix, so value of $\mu = [3.0138, -10.0188]$

5. Under assumption (iv), plot the log likelihood of observing this data-set as a function of μ where each component of μ belongs to $\{-10, ..., 10\}$ Compare this graph with the log-likelihood value you obtained in 3. What can you conclude?
   - Plot of Loglikelihood function with respect to mean:

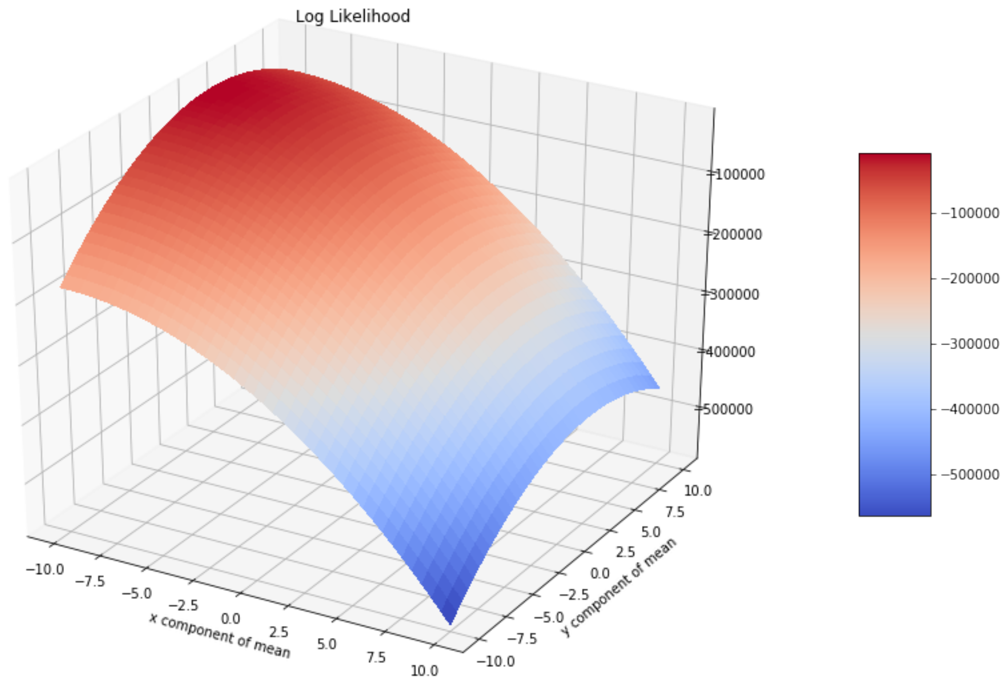


Figure 1.2: LogLikelihood plot

After plotting we find that maximum likelihood attend the maximum at same mean because it is independent of co-variance matrix. In the case where co-variance matrix is assumed [[1,0],[0,1]]. we find MLE for the same is somewhat less. So according to maximum likelihood estimate this can not be our best model for the given data.

Chapter 2

Gaussian Mixture Model

1. For every choice k, run your code to estimate the parameters of the model. Tabulate all the values you obtain.

mean	
NO. OF CLUSTERS	MEANS OF CLUSTERS
1	[-1.0328901900000003]
2	[-5.083402110538865, 1.3913407250224654]
3	[-5.007230212568095, -0.20036795174229444, 5.1276420884679395]
4	[-0.27709836531123433, -5.0087160314240515, -0.04942734402410028, 5.125839816698181]
5	[-5.00114133816751, -0.36366193875282704, 0.22249703661955234, 0.812234858557956, 5.239025489153628]
6	[-4.400575702558128, -5.698584385763029, -0.4416778893215279, -3.6811258891131353, 1.2520341841865406, 5.244393493218767]
7	[-5.004309391658733, -0.4032249254430587, -1.0842016647274029, -0.2803689081602233, 0.0018355050484639356, 2.24175599450678, 5.213563231985023]
8	[-5.741078960367857, -0.05268429390380408, -4.317308386304943, -1.052591357084585, -0.839985722811552, 0.056574943234379796, -0.28442615055450793, 5.190254871397996]
9	[-5.049594274442769, -0.2848356522265299, -2.6149109392884973, -0.23736113332003428, 0.13372886001854034, 0.44376646670625386, 0.5620175833330358, 5.255263350727185, 1.2807181153797962]
10	[-5.72659126527888, -3.849984788193701, -1.4299529313800505, -4.394226547732695, -0.9892117485180278, 0.3965187658613741, 0.44242106131077663, 0.4210455795132948, 1.8987972980399597, 5.246824736954268]

variance	
NO. OF CLUSTERS	VARIANCE OF CLUSTERS
1	[17.010698069063693]
2	[0.9520777940290989, 10.92550869382575]
3	[1.0588261501905818, 2.367024556154853, 1.2112741936254592]
4	[2.4937170696399535, 1.0573344903032385, 2.079492560331563, 1.2140927369911756]
5	[1.0746435473291363, 1.4258250115940816, 5.251346847117363, 5.007328126065672, 1.077748140848836]
6	[0.09439735465078668, 0.4829786328297486, 1.1977657873484568, 0.7515961467866238, 3.0312996030357895, 1.0706214910266407]
7	[1.0695168211564652, 1.6718240943145262, 3.8560183263920833, 1.2312173795199854, 2.676928204716838, 2.797849159347179, 1.111939161989322]
8	[0.458279818320381, 1.1657068214411246, 0.24983358554409688, 4.525693796428465, 4.786623343385607, 5.4079541477174295, 5.300494662416809, 1.1394733153489376]
9	[1.0213097294877662, 1.0412526556633415, 2.5399387064997305, 4.43360660706788, 4.633940232600413, 4.58701174001217, 4.530864008546987, 1.0584569241862818, 4.022221240594095]
10	[0.4728801330930925, 1.0891153741698958, 3.314543082182774, 0.09335632493328307, 0.10004353677025493, 0.5901861752228886, 4.753614609622795, 4.758501809250881, 3.320827160778065, 1.0740470525655854]

mixing coefficient	
NO. OF CLUSTERS	**mixing coefficient** OF CLUSTERS
1	[1.0]
2	[0.3744134673129909, 0.6255865326870086]
3	[0.4225477492930408, 0.3524889750805865, 0.22496327562637264]
4	[0.2452913165396934, 0.42202366494009785, 0.10751402159656136, 0.22517099692364786]
5	[0.4202738162610688, 0.19193367477968526, 0.1358504651484454, 0.04775459124655617, 0.20418745256424461]
6	[0.08954320900130802, 0.23598215527347224, 0.21476452852055067, 0.12184987273120738, 0.13088873141586654, 0.20697150305759487]
7	[0.4203702641768564, 0.09832560598912117, 0.05512261555181751, 0.07341007888921496, 0.10037576828389601, 0.04216258283902277, 0.21023308427007065]
8	[0.21914319296050952, 0.11992236459054885, 0.16125175461581137, 0.11127714584117507, 0.045246342822315294, 0.03456056811851381, 0.09512981710769255, 0.21346881394343337]
9	[0.40263236720670287, 0.1456707974159636, 0.043624625488502766, 0.042078123180504995, 0.04988391810534885, 0.022648920008937732, 0.04751894848043856, 0.20313855346087004, 0.04280374665273096]
10	[0.21957234853538693, 0.12796794065758765, 0.05318631182487391, 0.08591618320617242, 0.059401973923387155, 0.09902880282627957, 0.037625188475432995, 0.07457415159566673, 0.03922395101349388, 0.2035031479417184]

2. Plot the log-likelihood of the data-set for the parameters estimated as a function of the number of mixtures k. What can you conclude about the process that generated this data-set using this plot?

- Plot of Log likelihood function with respect to number of clusters:

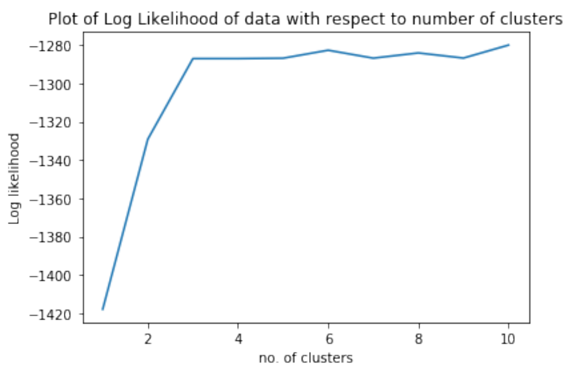


Figure 2.1: Log Likelihood plot of data with respect to number of clusters

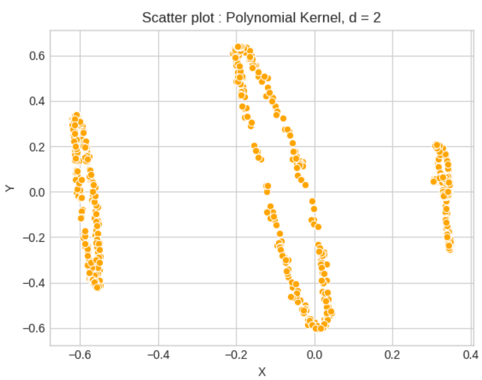
From the graph it can be seen that as clusters are increasing function is keep on increasing. But we can see, till 3 clusters it increases sharply and after that not much improvement.So, we can say best number of clusters is 3 and after that overfit happens.

Chapter 3

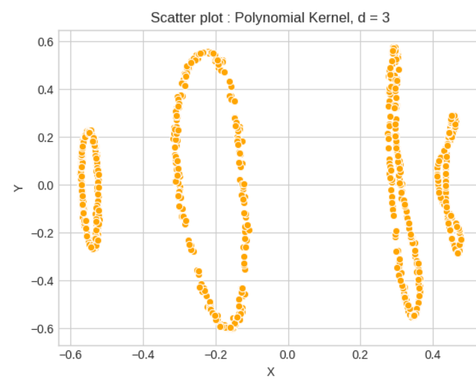
PCA & KPCA

1. How much of the variance in the data-set is explained by each of the principal components?
   - After applying PCA, each eigen value is explains variance associated to each principal component.
   - Eigen values : $\lambda_1 = 17.149, \lambda_2 = 14.504$
   - Variance by each of the principal components is given by $\frac{\text{Eigen value}}{\text{sum(all Eigen values)}} * 100\%$
   - The variance explained by first principal component with highest eigen value is **54.178**% and for second component value is **45.822**%. Cumulative sum is **54.178**% + **45.822**% = **100**%

2. Plot the projection of each point in the dataset onto the top-2 components for each kernel. Use one plot for each kernel and in the case of (B), use a different plot for each value of σ.
   - Plots for Kernel 1 :

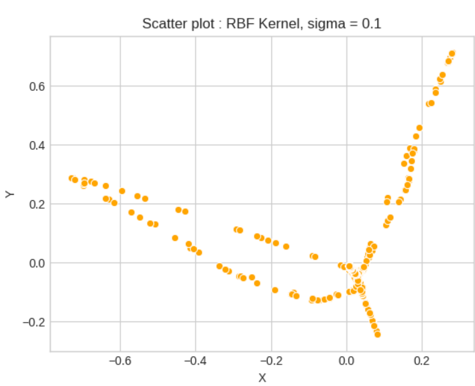


(a) Polynomial Kernel, d=2

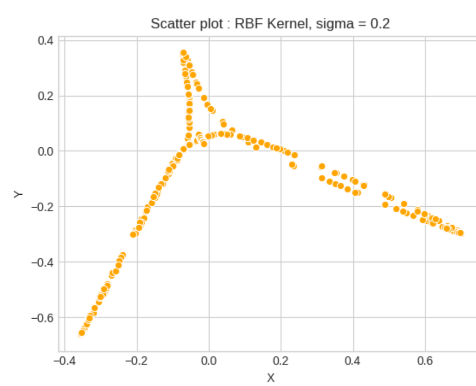


(b) Linear Kernel, d=3

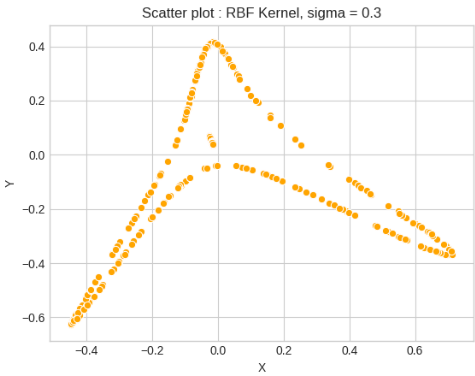
   - Plots for Kernel 2 :



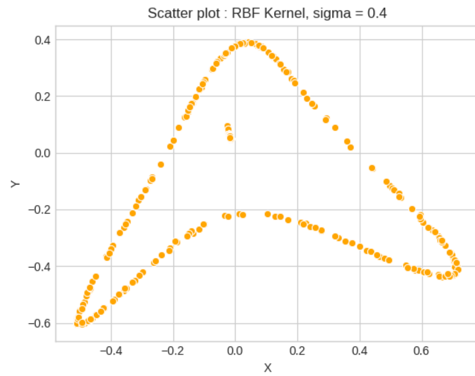
(a) RBF Kernel,$\sigma = 0.1$



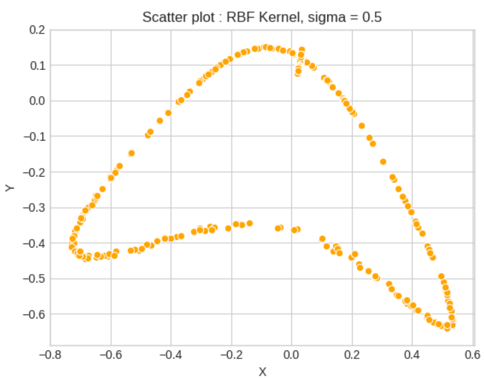
(b) RBF Kernel,$\sigma = 0.2$



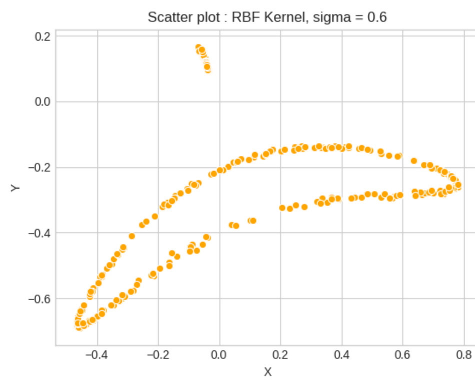
(a) RBF Kernel,$\sigma = 0.3$



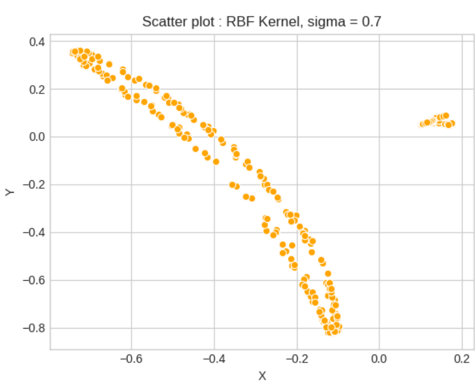
(b) RBF Kernel,$\sigma = 0.4$



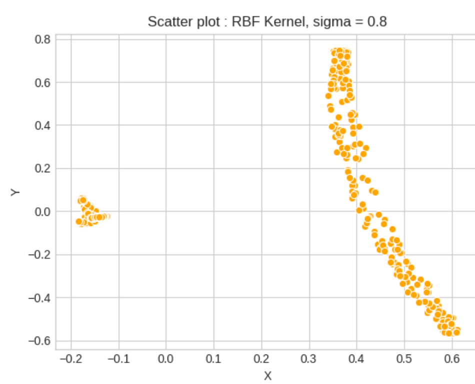
(a) RBF Kernel,$\sigma = 0.5$



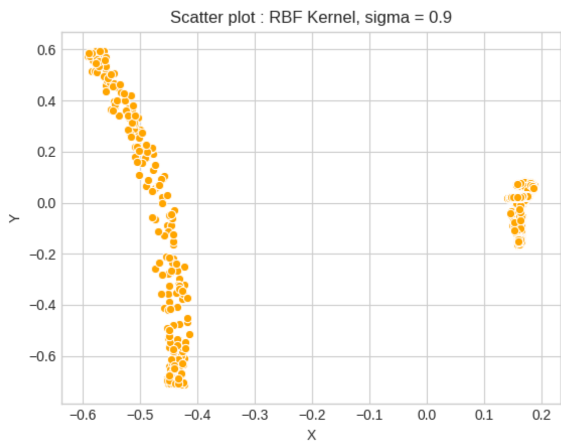
(b) RBF Kernel,$\sigma = 0.6$



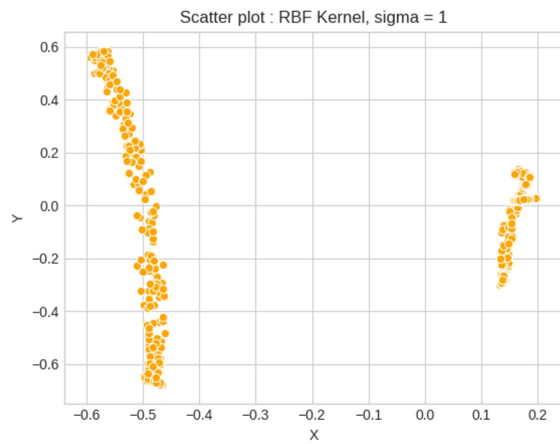
(a) RBF Kernel,$\sigma = 0.7$



(b) RBF Kernel,$\sigma = 0.8$

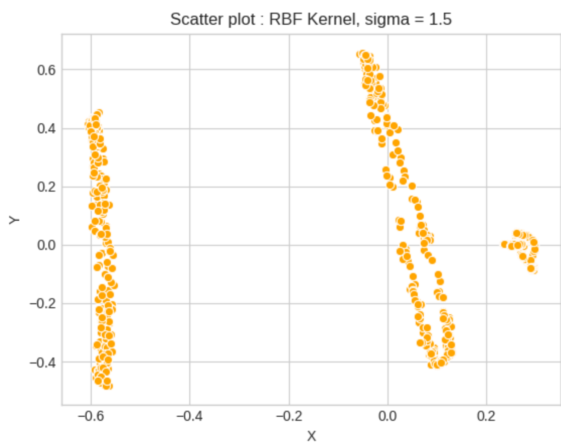


(a) RBF Kernel,$\sigma = 0.9$

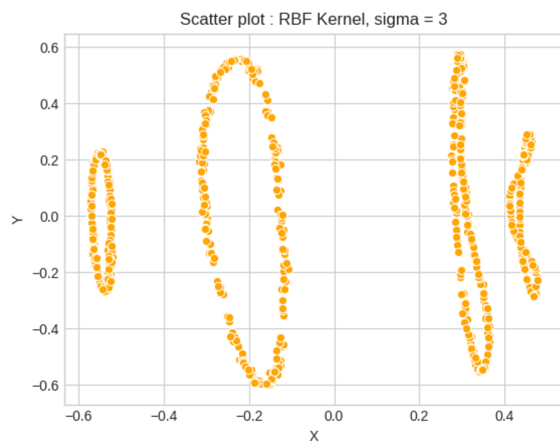


(b) RBF Kernel,$\sigma = 1$

If We increase further the value of σ, then we can able to classify given dataset into 4 classes as in figure3.1b.



(a) RBF Kernel,$\sigma = 1.5$



(b) RBF Kernel,$\sigma = 3$

3. Which Kernel do you think is best suited for this data set and why?

   - Both Polynomial and RBF kernel is giving good result on given data set.
   - **Polynomial kernel** is best in terms of computation and complexity. We shouldn't select complex model if simple is giving same result! As data set grows the kernel computation complexity increase significantly than Polynomial function!

Bibliography

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[2] C. S. Ong M. P. Deisenroth, A. A. Faisal. *Mathematics for Machine Learning*. Cambridge University, 2019.