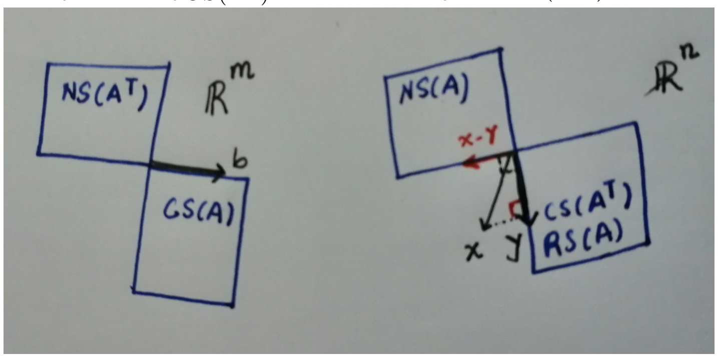
1. (5 points) The column space of a matrix $A \in R^{d \times d}$ is the set of all vectors that can be obtained as linear combinations of the column vectors of A. The row space of A is the column space of A^T.

   (a) Is the column space of a matrix $A \in R^{d \times d}$ same as the columnn space of $AA^\top$? If yes, prove. If not, give a counter example.

Solution:
True, Because
Column Space of $AA^\top \subset$ Column Space of A, The reason is Columns of $AA^\top$ are linear combinations of A itself. So any vector spanned by $AA^\top$ will be spanned by A.
Let $b \in CS(A)$ Therefore, there exist an x such that $Ax = b$
Let $y = Proj_{CS(A^\top)}x$ Therefore $y \in CS(A^\top)$ and $x - y \perp CS(A^\top)$



Let $y = A^\top u$, we claim : $AA^\top u = b$
$AA^\top u = Ay$
$x - y \perp ColumnSpace(A^\top) = RowSpace(A) \Rightarrow A(x - y) = 0 \Rightarrow Ax = Ay = b$
Therefore, $AA^\top u = b$, Column Space of $A \subset$ Column Space of $AA^\top$
Therefore, column space of a matrix $A \in R^{d \times d}$ same as the columnn space of $AA^\top$

   (b) Is row space of A same as column space of A? If so, prove. If not, argue why.

Solution: No, Row space of A is not same as column space of A.
Column space of A is vector space spanned by the columns of A. All linear combination of columns of A.
For Eg: Let A be 3 x 3 $\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$
The Column vectors are independent.
The Column space of A is span of $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$

Row space of A is the vector space generated by all linear combination of rows of A.

The Row space of A is span of $\begin{bmatrix}1\\4\\7\end{bmatrix}, \begin{bmatrix}2\\5\\8\end{bmatrix}, \begin{bmatrix}3\\6\\9\end{bmatrix}$

The vector $\begin{bmatrix}3\\9\\15\end{bmatrix}$ is spanned by basis of Row space but not by basis of Column space of A.

This shows that the Row space of A is not same as Column space of A.

If A is m x n Matrix then the Row Space and Column Space are of different dimensions.

Row Space $\in R^n$ and Column Space $\in R^m$

The only matrix that has Same Row Space and Column Space is a Symmetric Matrix where Rows are same as column.

1. (5 points) Let K1,K2 be two arbitrary kernel functions mapping vectors from $R^d \times R^d \Rightarrow R$. For each of the cases below, show if it is a valid Kernel or if it is not, argue why.

   (a) K3(x, y) = K1(x, y) + K2(x, y) + 7.5

   Solution:
   K3(x,y) is a valid kernal
   K = K1 + K2 + c (Closed under addition)
   We add matrices element wise:

   $$\forall x, X^\top KX = X^\top K1X + X^\top K2X + c \geq 0$$

   This follows the Mercer's Theorem property.
   Addition of elements will not affect the λ ,K3 will be PSD that is $\forall\lambda \geq 0$
   Hence, a valid kernel.

   (b) K4(x, y) = 5 K1(x, y) - 3 K2(x, y)

   Solution:
   No, Not a valid kernel.
   $K(x, y) = \alpha K1(x, y) + \beta K2(x, y), for\ \alpha, \beta \geq 0$
   This property is voilated as β<0
   Mercer's Theorem can be voilated, So K4(x,y) won't be PSD.

   (c) K5(x, y) = K1(x, y) * K2(x, y)

Solution:
K5(x,y) is a valid kernal
K = K1 * K2 (Closed under multiplication)
We multiply matrices element wise:

$$\forall x, X^\top K X = X^\top K1 X * X^\top K2 X \geq 0$$

This follows the Mercer's Theorem property.
Multilication of elements will not affect the λ ,K3 will be PSD that is $\forall \lambda \geq 0$

(d) K6(x, y) = $(x^\top y + 1)^3$

Solution:
This is a Polynomial Kernel.
$(x^\top y + 1)^3 = (x^\top y)^3 + 1^3 + 3(x^\top y) * 1 * (x^\top y + 1)$
$x, y \in R^d$

$(\sum_i^d x_i y_i + 1)(\sum_j^d x_j y_j + 1)(\sum_z^d x_z y_z + 1)$

$(\sum_i^d \sum_j^d \sum_z^d x_i y_i x_j y_j x_z y_z + \sum_i^d \sum_j^d x_j y_j x_i y_i + \sum_j^d \sum_z^d x_z y_z x_j y_j + \sum_z^d \sum_i^d x_i y_i x_z y_z + \sum_i^d x_i y_i + \sum_j^d x_j y_j + \sum_z^d x_z y_z + 1)$
All these forms new features, These is valid kernel as kernal is closed under addition and multiplication. Also follows Mercer's property, K6 is P.S.D.

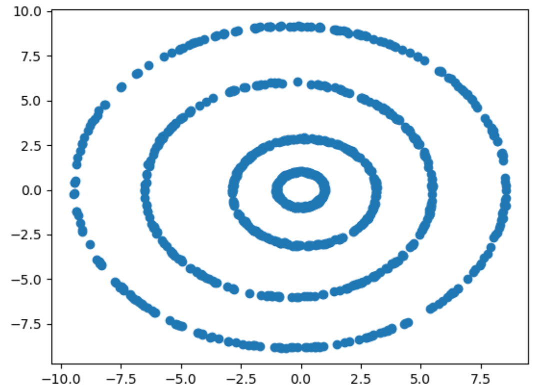
1. (5 points) You are given a data-set with 1000 data points each in R^2

(a) Write a piece of code to run the PCA algorithm on this data-set. How much of the variance in the data-set is explained by each of the principal components?

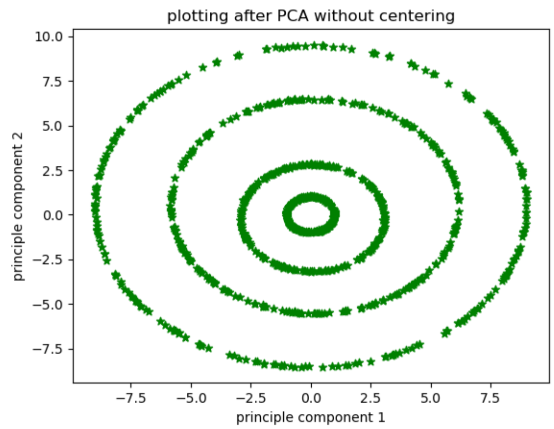
Solution:
Variance from first principle component 54.17802452885226
Variance from second principle component 45.82197547114777

(b) Study the effect of running PCA without centering the data-set. What are your observations? Does Centering help?

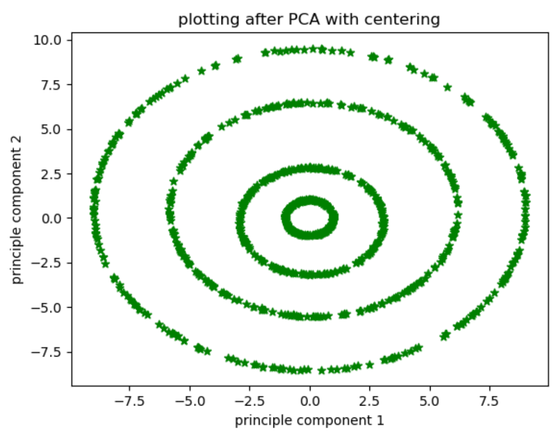
Solution:
Centering has no effect on the data.



Original data



Without centering



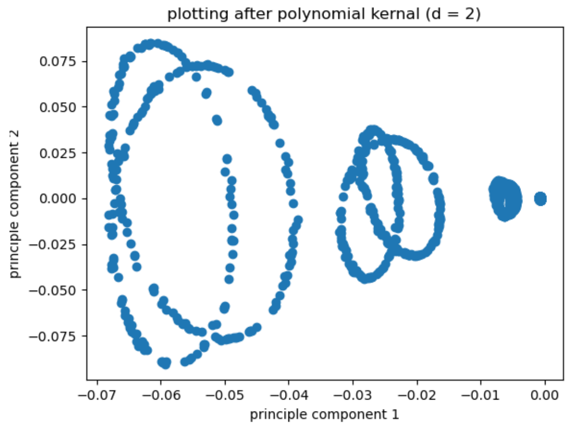
with centering

(c) Write a piece of code to implement the Kernel PCA algorithm on this dataset.Use the following kernels : A. (x, y) = (1 + xT y)d for d = 2, 3 B. (x, y) = exp (xy)T (xy)/22 for = 0.1, 0.2, . . . , 1

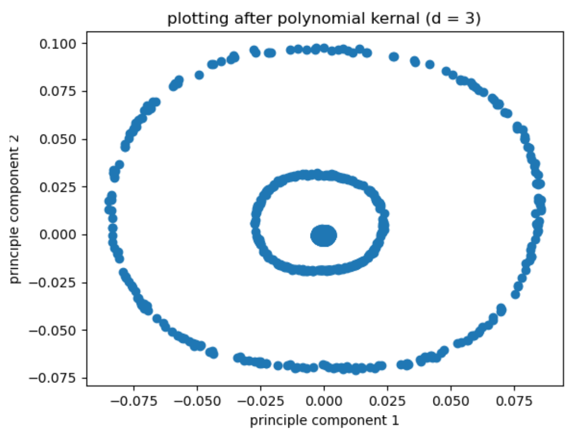
Plot the projection of each point in the dataset onto the top-2 components for each kernel. Use one plot for each kernel and in the case of (B), use a different plot for each value of .

Solution:

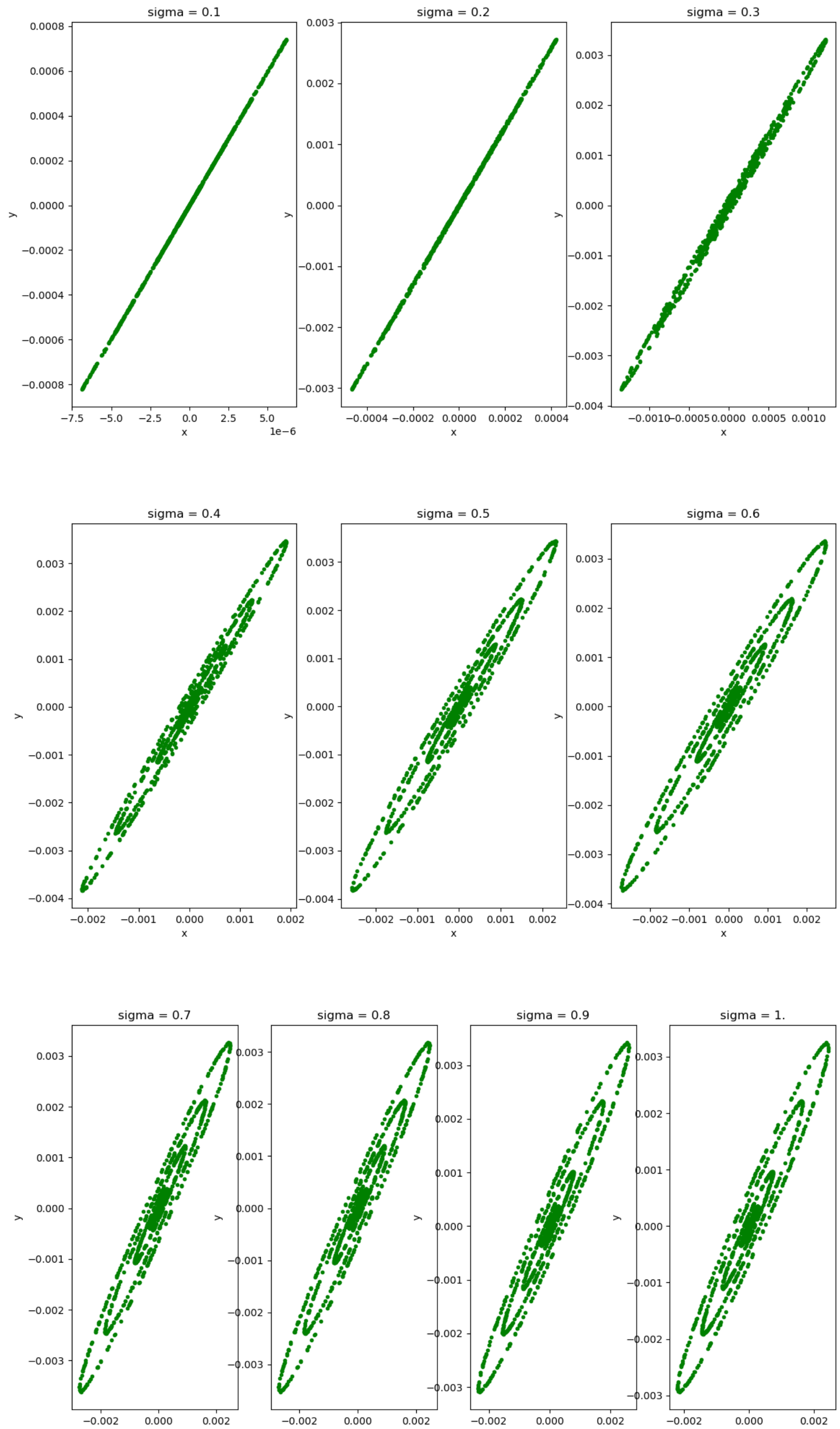
Polynomial kernal d=2



Polynomial kernal d =3



Gauusian kernal


sigma = 0.1
sigma = 0.2
sigma = 0.3
sigma = 0.4
sigma = 0.5
sigma = 0.6
sigma = 0.7
sigma = 0.8
sigma = 0.9
sigma = 1.

(d) iv. Which Kernel do you think is best suited for this dataset and why?

Solution:
Polynomial kernal with d = 2 find the better clusters in the data.

Play Question: Non-Mandatory Create your own image dataset as follows. Use your phone to capture 25 photos of some type of vessel (category 1) in your kitchen and 25 photos of chairs/tables in your home (category 2). (If these are not available, pick any two sets of distinct items. You can take multiple pictures of the same item with slightly different angles). Pick only 20 images in each category and convert them to black and white. Make sure they are of the same size. Let the pixel intensity be your features. Run standard PCA and project all the 40 images onto the Eigenspace spanned by the top Kcategory 2. What do you observe from this experiment? Draw insights .

Solution:

Avg. distance from category1 : [[3142.03045837 3156.92568704 3112.56052044 3123.37168525 3166.31177588 2034.97185195 1932.6797632 1963.20140869 2009.61054991 1950.8610057] [3150.8031543 3169.01497839 3141.39283706 3153.12381703 3177.06994618 2044.759875 1961.0156153 1991.62245422 2046.14905941 1979.13599293] [3159.54291188 3181.81535828 3170.50542434 3184.36495604 3188.31796156 2052.37255299 1985.00717546 2014.05057835 2084.08992874 2005.00852946] [3174.93687203 3201.74883963 3218.25841966 3233.40423744 3207.29326127 2067.94848349 2031.34156996 2056.87131037 2149.9094585 2050.90488732] [3196.12245333 3230.08957496 3277.96482973 3296.25879218 3233.53548108 2087.97688456 2084.62330534 2109.63323176 2221.70862252 2101.62622836] [3219.96952247 3263.99342807 3348.58205642 3371.2269343 3259.01832662 2106.14553798 2144.0137845 2163.31530307 2289.48906636 2161.63654906]] Avg. distance from category1 : [[1381.22658088 1510.69249013 1671.92621176 1679.7763668 1537.60027579 3049.46682232 2977.32665985

3006.26598152 2964.13251319 2942.88379637] [1402.22829132 1536.75172597 1727.07097077 1736.44844347 1561.89437237 3055.04100343 2994.69470581 3023.96280356 2988.03140405 2960.50688906] [1422.814889 1563.90789782 1781.2414024 1794.30476338 1586.80886789 3059.4449969 3009.58280769 3038.07934152 3013.17928971 2976.8457497] [1458.330087 1605.371897 1867.26885865 1882.18748582 1627.83110237 3068.61428888 3038.78462287 3065.38619774 3057.56630951 3006.29826304] [1505.81810398 1662.75874478 1970.6672635 1990.42310367 1682.74180883 3080.66608729 3073.05144085 3099.64165986 3107.01967857 3039.49515486] [1557.59108927 1729.27944795 2088.06011111 2114.22276642 1734.32572299 3091.80998584 3112.05179483 3135.14474008 3154.62824797 3079.58185791]] [['Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat1', 'Cat1', 'Cat1', 'Cat1', 'Cat1'], ['Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat1', 'Cat1', 'Cat1', 'Cat1', 'Cat1'], ['Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat1', 'Cat1', 'Cat1', 'Cat1', 'Cat1'], ['Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat1', 'Cat1', 'Cat1', 'Cat1', 'Cat1'], ['Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat1', 'Cat1', 'Cat1', 'Cat1', 'Cat1'], ['Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat2', 'Cat1', 'Cat1', 'Cat1', 'Cat1', 'Cat1']]

The algorithm identifies the two different categories correctly.