1. (a) **Solution:** given A as dxd matrix, so no of columns no of rows.

   we also know that,
   rank(A)=rank($A^\top$)= no of columns(A) - dimension of null space of (A)
   = no of columns($A^\top$) - dimension of null space of ($A^\top$)
   = no of columns of (A) - dimension of null space of ($A^\top$) (as A is dxd matrix)
   from above we can say that,
   dim(null space(A)) = dim(nullspace($A^\top$)) -equation(1)

   we know that, null space of ($AA^\top$) = null space of $A^\top$(i.e if $A(A^\top v) = 0 = A^\top v$)
   so, dim(nullspace($AA^\top$)) = dim(nullspace($A^\top$))
   now, form equation(1) we have,
   dim(nullspace($AA^\top$)) = dim(nullspace($A^\top$))=dim(nullspace(A))
   rank($AA^\top$)= no of columns of $AA^\top$ - dim(null space of $AA^\top$)= dim(null space of A)
   as no of columns A =no of columns of $AA^\top$
   we have , rank(A)= rank($AA^\top$)
   so, no of linearly independent columns of $AA^\top$ = no of linearly independent columns of A
   dim(column space($AA^\top$)) = dim(column space(A)) -(2)
   we know that, column space of $AA^\top \subseteq$ column space of A, as columns of $AA^\top$ will be linear combination of columns of A and dim(column space($AA^\top$)) $\leq$ dim(column space(A)) -(3)
   hence, from (2) and (3) we can say that column space of A is same as column space of $AA^\top$

   (b) **Solution:** no, row space of A may not be same as column space of $A^\top$ ex:
   let A= $\begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$
   here column space basis=$\{[1\ 1\ 1]^\top\}$ and row space basis= $\{[1\ \ -1\ \ -1]^\top\}$,so both row space and column space are different here.

2. considering k1(x,y) and k2(x,y) as valid kernel mapping,
   for a kernel function to be valid it has to be symmetric and positive semi definite matrix.
   as k1 and k2 are valid kernel so k1(y,x)=k1(x,y) and k2(y,x)=k2(x,y) – (1)
   let matrix obtained by kernel,k2 is K2, and k1 is K1
   as k1 and k2 are valid kernels so, K1 and K2 will be positive semi definite matrix. – (2)

if $c^\top$Kc ≥ 0, $\forall c \in R^n$ then it will be positive semi definite matrix. considering K as nxn matrix.

Solution: given , k3(x,y)=k1(x,y)+k2(x,y)+7.5
for symmetric matrix,
k3(y,x)=k1(y,x)+k2(y,x)+7.5
k3(y,x)=k1(y,x)+k2(y,x)+7.5=k3(x,y) (from (1))
so k3 will create symmetric matrix.
for, positive semi definite

let matrix obtained by kernel, k3 is K3
now, $c^\top K3c = c^\top(K1 + K2 + 7.5I)c$ $\forall c \in R^n$ here I is identity matrix
$c^\top K3c = c^\top K1c + c^\top K2c + c^\top 7.5Ic)c$ we know that,I is also a positive semi definite matrix also 7.5¿0.
$c^\top K1c \geq 0$, $c^\top K2c \geq 0$ and $c^\top 7.5Ic \geq 0$ so, we have $c^\top K3\ c \geq 0$ hence k3(x,y) is a valid kernel function

(b) **Solution:** given , k4(x,y)=5 * k1(x,y)- 3 * k2(x,y)
for symmetric matrix,

k4(y,x)=5 * k1(y,x)- 3 * k2(y,x)
k4(y,x)= 5 * k1(x,y) - 3 * k2(x,y) (from (1))
k4(y,x)=k4(x,y) so k4 is symmetric kernel

let matrix obtained by kernel k4 is K4
now, $c^\top K4c = c^\top(5 * K1 - 3 * K2)c$, $\forall c \in R^n$
$c^\top K4c = 5 * c^\top K1c - 3 * c^\top K2c$

$c^\top K1c \geq 0$ and $c^\top K2c \geq 0$ (form (2))

but, for -3 * $c^\top$ K2 c ≤ 0
so, we can't say anything about K4 it may or may not be positive definite w.r.t to given data points.
hence k4(x,y) is not always a valid kernel function

(c) **Solution:** given , k5(x,y)= k1(x,y) * k2(x,y)
for symmetric matrix,

k5(y,x)= k1(y,x) * k2(y,x)
k5(y,x)= k1(x,y) * k2(x,y) (from (1))
k5(y,x)=k5(x,y) so k5 is symmetric kernel

as k1 and k2 are valid kernel they can be represented as inner product due to mercers theorem. let a be the feature vector for k1 and b for k2. k1(x,y)=a(x)$^\top$a(y) , k2(x,y)=b(x)$^\top$b(y) where,
a(k)=$[a_1(k), a_2(k), \ldots, a_m(k)]$ and b(k)=$[b_1(k), b_2(k), \ldots, b_n(k)]$

k5(x,y)=k1(x,y) * k2(x,y)
= a(x)$^\top$a(y) * b(x)$^\top$b(y)
=$\sum_{i=1}^{i=m} a_i(x)a_i(y) * \sum_{j=1}^{j=n} b_i(x)b_i(y)$
=$\sum_{i=1}^{i=m} \sum_{j=1}^{i=n}[a_i(x)b_j(x)] * [a_i(y)b_j(y)]$
=$\sum_{i=1}^{i=m} \sum_{j=1}^{i=n} c_{ij}(x)c_{ij}(y)$
=$c(x)^\top c(y)$
here k5 can be represented as inner product using feature map c, so due to mercers theorem we can say that k5 is a positive semidefinite matrix. hence k5(x,y) is a valid kernel function

(d) **Solution:** given, k6(x,y)=$(x^\top y + 1)^3$
k6(y,x)=$(y^\top x + 1)^3$
k6(x,y)=k6(y,x) $(y^\top xisscalar)$

so k6 is symmetric kernel

we know that simple covariance matrix i.e. A$^\top$A is positive semidefinite and symmetric
let k7 = x$^\top$y be the kernel function for this covariance matrix
let k8= x$^\top$y+1, from (a) we have k3=k1+k2+7.5 is valid kernel which says that k3=k1+c where c > 0 is also a valid kernel.
so k8 will be a valid kernel
let k9= k8 * k8,
from part (c) we have k5=k1*k2 is valid kernel
so k9 will be valid kernel
let,k10 = k9 * k8, so again from (c) we can say that k10 will be valid kernel

k10(x,y)= k9(x,y) * k8(x,y)= k8(x,y) * k8(x,y) * k8(x,y)
= (x$^\top$y+1) (x$^\top$y+1) (x$^\top$y+1)
=(x$^\top$y+1)^3 = $k6$

hence we have k6 as a valid kernel function

3. (a) **Solution:** the code for PCA is in attached code file named question 3 (part a). the total of variance observed by data along principal components is = 31.621519151774983
the variance along first principal component is 54.178025 % of total variance and along second principal component is 45.821975 % of total variance.

(b) **Solution:** by running the PCA on the given data without centering, the result obtained are as follows:
total of variance along principal components: 31.621519151774983
variance along first principal component: 54.178025 %
variance along second principal component: 45.821975 %
form (a) and (b) we can see that the percentage of variance observed by each of principal components are same in case of centered and non-centered data.
hence , we can say that for given data centering does not help
below figure is shown for the centered data

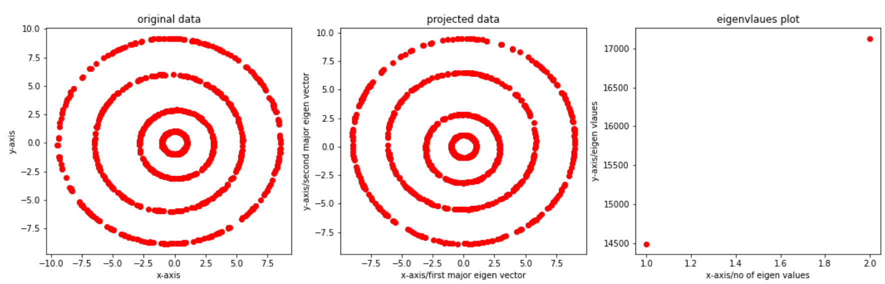


figure1:results of centered data
below figure shows the results for non-centered data

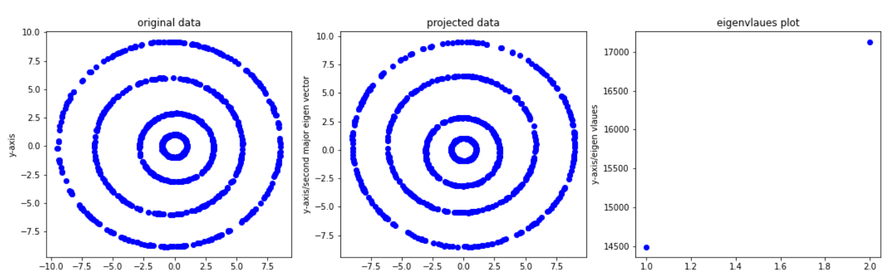


figure1:results of non centered data
as the given data is already centered so centering does not plays any role here. from figure it is clear that PCA is not able to do any helpful transformation on data, because, after projecting data onto eigen space we are obtaining the same data.

also we are having the initial total variance of given data = total variance of projected data.
as the data is circular around the origin, any pair of perpendicular vectors can become principal components directions.

(c) **Solution:** the plots for kernel k=$(x^{\top}y + 1)^d$ are shown below,

figure2: for d = 2

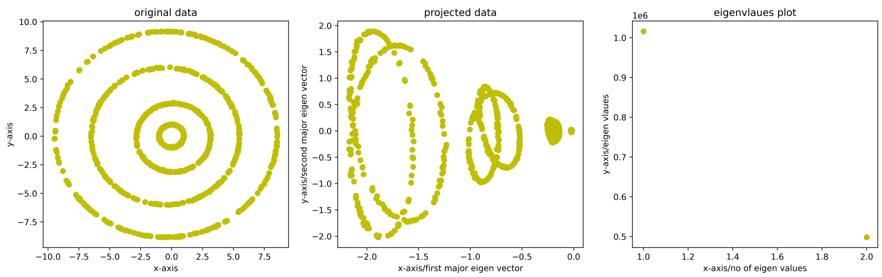
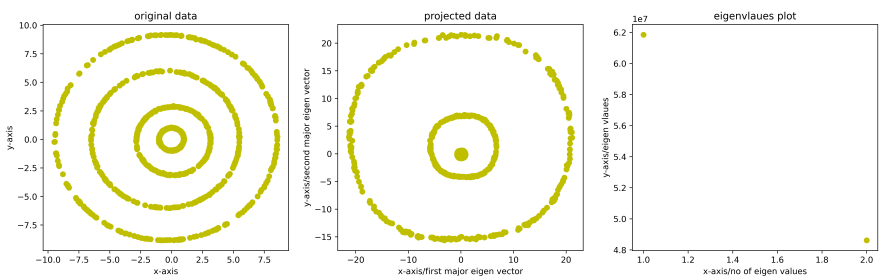


figure3: for d = 3



the plots for kernel k=exp $\frac{-(x-y)(x-y)^{\top}}{2\sigma^2}$ are shown below,

figure4: for σ= 0.1

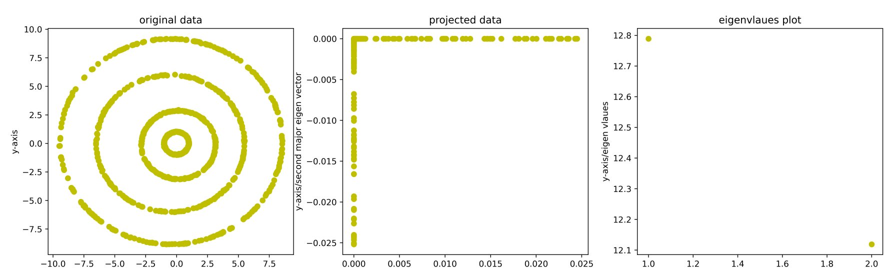


figure5: for σ= 0.2

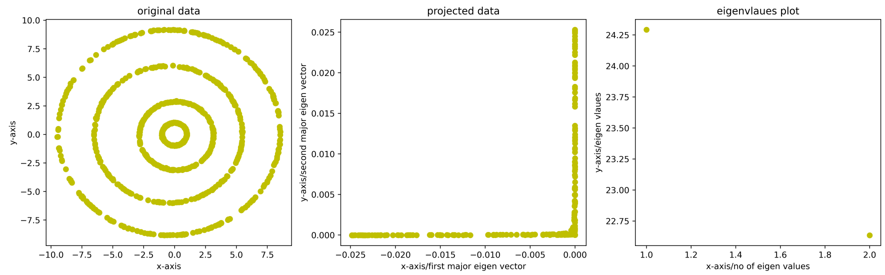


figure6: for σ= 0.3

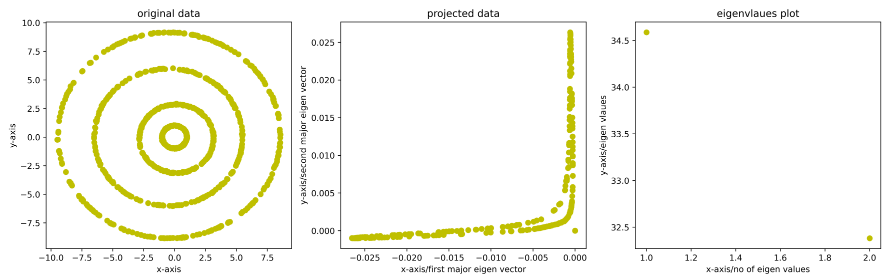


figure7: for σ= 0.4

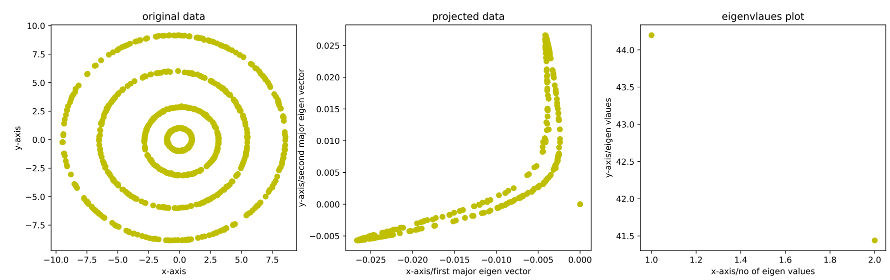


figure8: for σ= 0.5

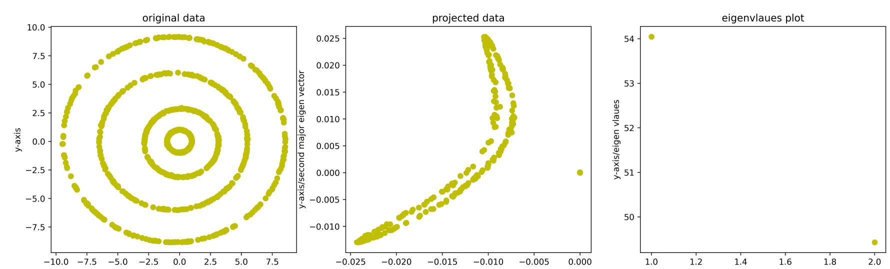


figure9: for σ= 0.6

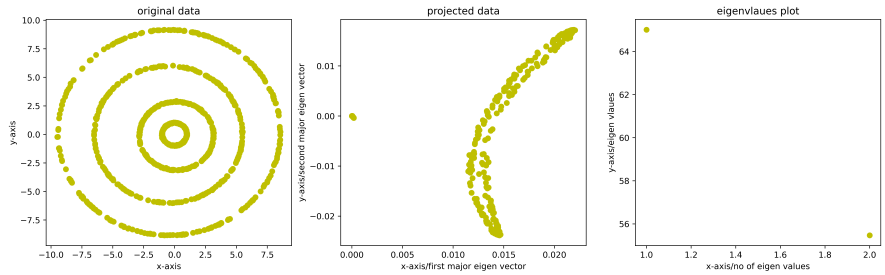


figure10: for $\sigma = 0.7$

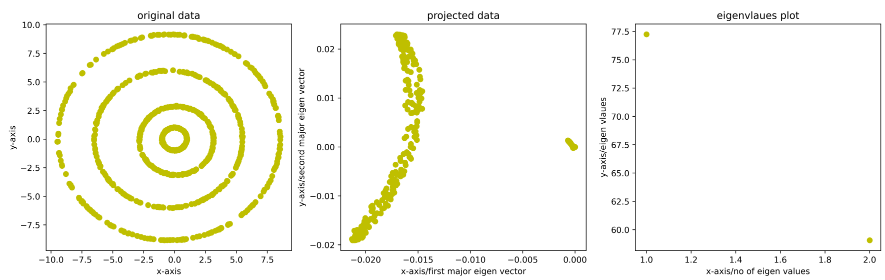


figure11: for $\sigma = 0.8$

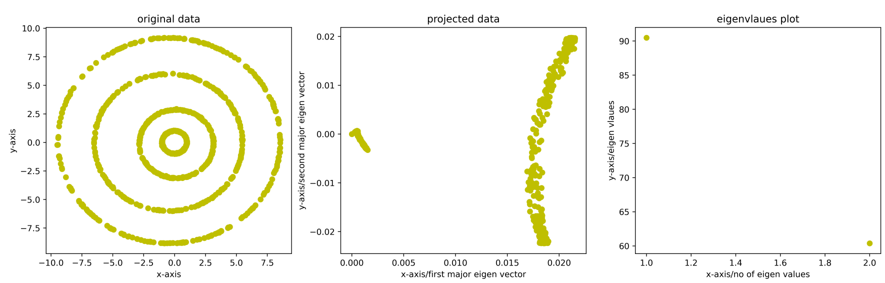


figure12: for $\sigma = 0.9$

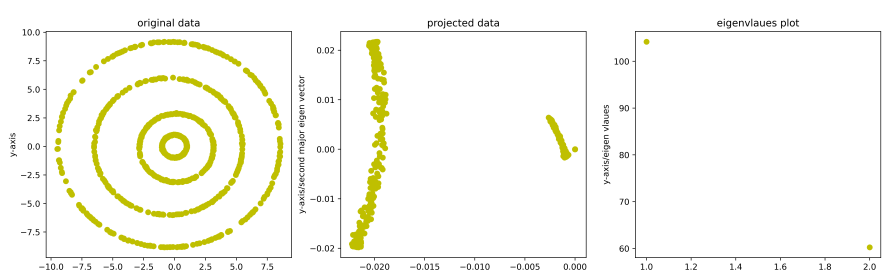
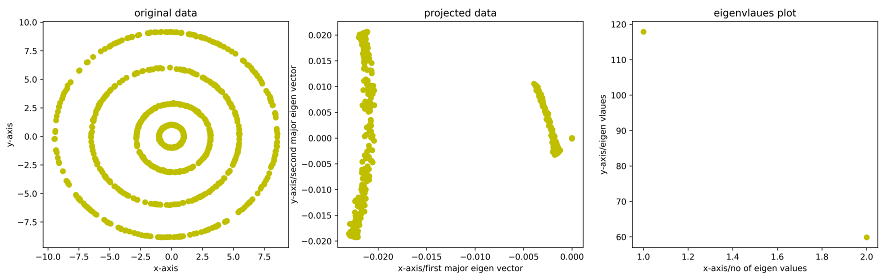


figure13: for $\sigma = 1$



(d) **Solution:** the gaussian kernel is best suited for this dataset with σ as 1,because it is capturing maximum variance among all other values of σ. the polynomial kernel is not able to properly decorelate the data it can be seen form figure 2 and figure 3 as compared to the gaussian kernel.

4. **Solution:**

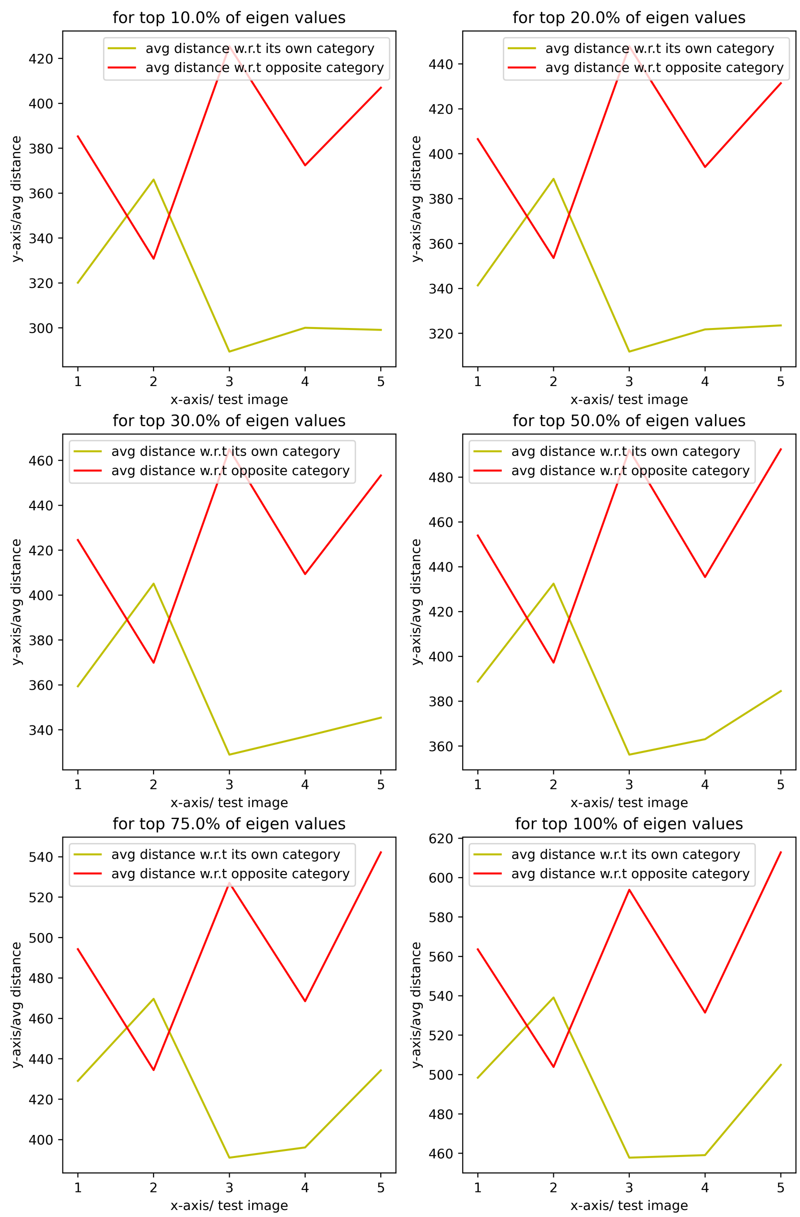


figure14: avg distance of category 1 test images w.r.t category 1 and category 2

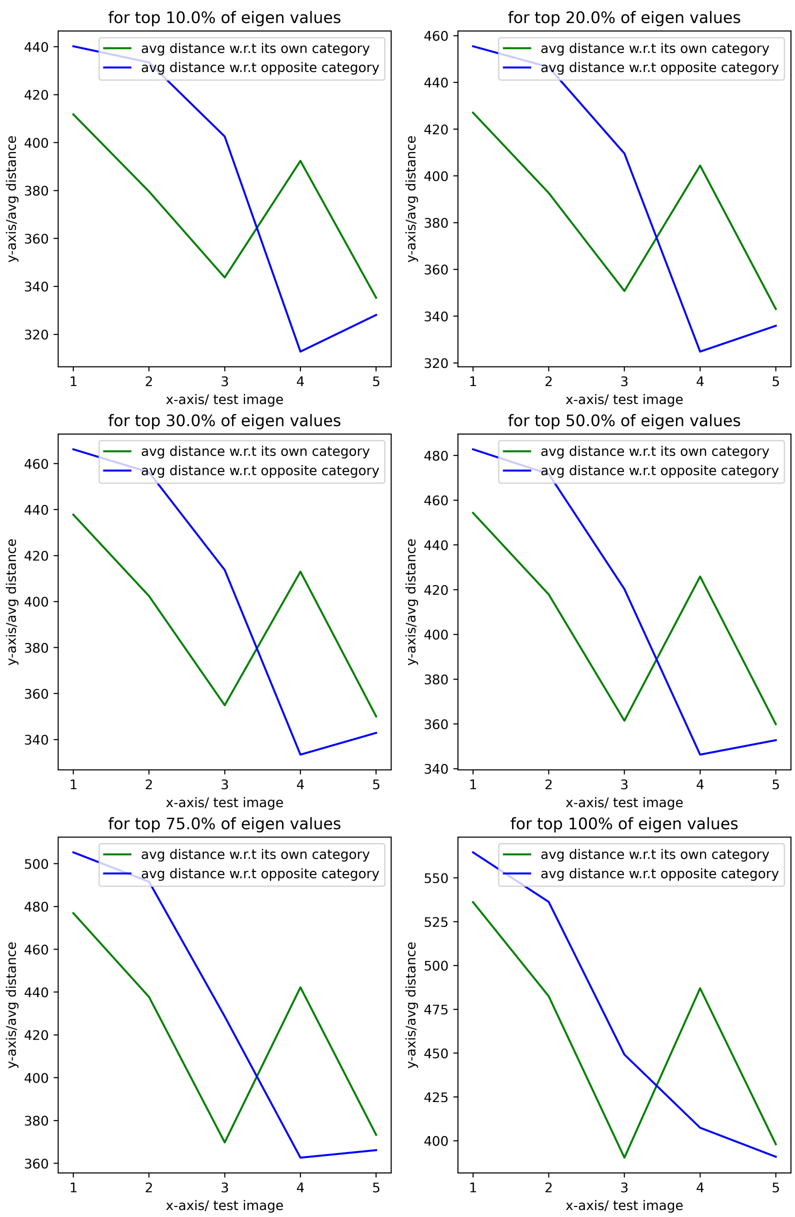


figure15:avg distance of category 2 test images w.r.t category 1 and category 2

from the above two figures we can see that, for most of the test image its average distance w.r.t to its own category is less as compared to its avg distance w.r.t. other category,and the difference between these two distance keep on decreasing as we increase the no of eigen vectors for most of test images.

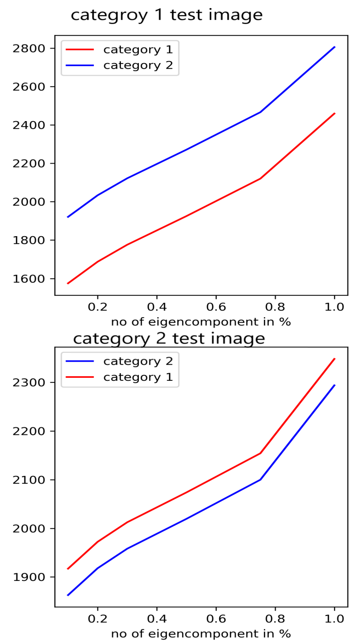


figure16:relation between no of eigen component and sum of avg distance of all test images of category 1, and category 2, w.r.t, category 1 and 2

from the above figure we can see that, the maximum difference for sum of average distance of category 1 test images w.r.t its own category images and w.r.t other category images is for top 10% to top 30% eigen components, and if we keep on increasing the eigen component this difference is decreasing. form this we can say that, on the basis of top 10% to top 30% eigen components we are highly able to say that the given set of test images belongs to which category.

code for this question is in file named question 4

all the code are in folder named code
dataset link: `https://drive.google.com/drive/folders/15cbU-9py4VnJ8jONj6umCr6l7XaTlOyT?usp=sharing`