Question 1:

First part:

We have to prove whether $C(A)=C(AA^T)$ is true or not:

Answer: Our Claim is that it is true

Proof: We know our if we multiply 2 matrices A with B then resultant will be just linear combination of column vectors of A. So, the column space of AB will be at maximum equal to the column space of A. Which is mathematically: **$C(A)>=C(AB)$**

Similarly we can say that **$C(A)>=C(AA^T)$**

Now, if we can prove **$C(A)<=C(AA^T)$** then it will be obvious that **$C(A)=C(AA^T)$**

To prove this we shall take help from Singular Value Decomposition.

Let $A = U\Sigma V^T$ [using Singular Value Decomposition where $U \in R^{n*d}$ and $\Sigma \in R^{d*d}$]

Σ is diagonal matrix and columns of U and V both are orthonormal…

So,

$AA^T . U\Sigma^{-1}V^T = . U\Sigma^2U^T . . U\Sigma^{-1}V^T$

$= U\Sigma^2 (U^T U)\Sigma^{-1}V^T$

$= U\Sigma^2 (I)\Sigma^{-1}V^T$ [As U is orthonormal so $U^T U=I$]

$= U\Sigma(\Sigma\Sigma^{-1})V^T$

$= U\Sigma V^T$

So, each column of A satisfies

$A = AA^T . U\Sigma^{-1}V^T$

So, each column of A is a linear combination of columns of AA^T

Which means column space of A will be at maximum Column space of AA^T

So, we can say that **$C(A)<=C(AA^T)$**

So, $C(A)>=C(AA^T)$

And **$C(A)<=C(AA^T)$**

So, $C(A)=C(AA^T)$ [proved]

Second Part:

We have to prove if row space of A in equal to column space of A or not.

Claim: Our claim is that it is false

Proof: We shall prove it by the help of a counter example:

Lets take a matrix A (2*2)=

1	2
9	18

Here Row space space is the span of vector $[1,2]^T$

RowSpace(A) = C * $[1,2]^T$ [C is the set of all real numbers]

Similarly Column Space is the span of vector $[1,9]^T$

ColumnSpace(A) = C * $[1,9]^T$ [C is the set of all real numbers]

Definitely Row space and column space are not equal

Question 2:

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

Let K1 = $f_1(x)^T f_1(y)$

And K2 = $f_2(x)^T f_2(y)$

[f_1 and f_2 are feature space of K1 and K2]

Now , K3 = K1 * K2 +7.5

We have to some how prove

K53 = $f_3(x)^T f_3(y)$

Where f_3 is the feature space of K3

Let us assume $f_3(x)$ =

$$\begin{bmatrix} f_1(x) \\ f_2(x) \\ \sqrt{7.5} \end{bmatrix}$$

and $f_3(y)$ =

$$\begin{bmatrix} f_1(y) \\ f_2(y) \\ \sqrt{7.5} \end{bmatrix}$$

Now

$f_3(x)^T f_3(y)$

$=[f_1(x)^T \;\; f_2(x)^T \;\; \sqrt{7.5}]$ *

$$\begin{bmatrix} f_1(y) \\ f_2(y) \\ \sqrt{7.5} \end{bmatrix}$$

$= f_1(x)^T f_1(y) + f_2(x)^T f_2(y) +7.5$

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

=K3

So, K3 can be expressed as a inner product and it is equal to

K1(x, y) + K2(x, y) + 7.5

So, K3 is a valid kernel accrording to mercer's theorem

K4(x, y) = 5 ∗ K1(x, y) − 3 ∗ K2(x, y)

We have to find whether this is valid or not

Claim: It is an invalid kernel

Proof by Counter Example:

Let us assume K1(x,y) = 2I [I=Identity matrix]

And K2(x,y)=5I

So, K4 = 5* K1 – 3* K2

= 5*2I – 3*5I

=10I-15I

=-5I

So we got K4= -5 * I

Defininetly eigenvalues of K4 are negative, so it is not in PSD

According Mercer's theorem it is not a valid kernel.

Hence it is an invalid kernel

K5(x, y) = K1(x, y) ∗ K2(x, y):

Claim: It is a valid kernel

Proof:

Let K1 = $f_1(x)^T f_1(y)$

And K2 = $f_2(x)^T f_2(y)$

[f_1 and f_2 are feature space of K1 and K2]

Now , K5 = K1 * K2

= $f_1(x)^T f_1(y) * f_2(x)^T f_2(y)$

Which is also feature space because it can be written as

= $(f_1(x)^T * f_2(x)^T) * f_1(y)\ f_2(y)$

$= (f_2(x)\ f_1(x))^{T} * (f_1(y)\ f_2(y))$

This is also an inner product,

So, K5 can be written as

$K5 = f_3(x)^T f_3(y)$

So, this is also a valid kernel according to mercer's theorem.

$K6(x, y) = (x^T y + 1)^3$

Let us assume $X = [x1\ \ x2]^T$ and $Y = [y1\ y2]^T$

Now $x^T y = x1y1+x2y2$

So, $x^T y$ is a valid kernel.

Now, lets see if $x^T y+1$ is a valid kernel or not

For that let us assume $f(x) = [x1\ x2\ 1]^T$ and $f(y) = [y1\ y2\ 1]^T$

So, $f(x)^T f(y) = x1y1+x2y2+1 = x^T y + 1$ (from last step)

So, $x^T y+1$ is a valid kernel

Let us assume $K6 = (K)^3$ [Where $k= x^T y+1$]

Again assume we are representing K as K(x',y')

Where x' = [x1' x2']

y' = [y1' y2']

Let $f_1(x') = [x1'^3 \quad x2'^3 \quad \sqrt{3}\,(x1')^2 * x2' \quad \sqrt{3}\,(x2')^2 * x1']$

$f_1(y') = [y1'^3 \quad y2'^3 \quad \sqrt{3}\,(y1')^2 * y2' \quad \sqrt{3}\,(y2')^2 * y1']$

Now, $f_1(x')^T * f_1(y') = x1'^3 y1'^3 + x2'^3\ y2'^3 + 3\ x1'^2 y1'^2\ x2' y2' + 3\ x2'^2 y2'^2\ x1' y1'$

$= (x1y1+x2y2)^3$

$=K^3$

So, K is also a valid Kernal according mercer's therom

So, **K6 is a valid kernal**

Question 3:

i.

PCA algorithm has been run on the dataset.The program file is - simple pca with centering.ipnyb

Variance of first principal component: 17.1319144024 (var_of_component1)

Variance of second principal component: 14.4896047493 (var_of_component2)

Variance explained by first principal component: 54.1780245289% (percent1) of total variance

Variance explained by second principal component: 45.8219754711% (percent2) of total variance

ii.

PCA alrogithm has been run on the dataset without centering.The program file name is :

simple pca without centering.ipnyb .

The result is same,

i.e.

Variance of first principal component: 17.1319144024 (var_of_component1)

Variance of second principal component: 14.4896047493 (var_of_component2)

Variance explained by first principal component: 54.1780245289% (percent1) of total variance

Variance explained by second principal component: 45.8219754711% (percent2) of total variance

This is because the data provided is already centered.we can see that in the first program file (simple pca with centering.ipnyb) also.

So, Centering didn’t help

iii.

Program name for Kernel PCA on dataset A is: Kernal PCA assignemnt on A

Program name for Kernel PCA on dataset B is: Kernal PCA assignemnt on B

iv.

The kernel B is best suitable.This is because-

Running kernel PCA on A is not fulfilling our goal, that is our data is still not linearly distinguishable, we cannot find any linear relation both for d value 1 and 2

But, while running this on B, specially when we increase the value of sigma, we can see that even though the data was not linearly distinguishable but when sigma was increased at value 0.9 and 1 , we found out too some extent linear relation on our Dataset. That means we were able to separate data linearly.

So, kernel B is best suitable.