Question 5 - kernel PCA

1. The kernel matrix is defined as $K_{i,j} = \langle \phi(x_i), \phi(x_j) \rangle$
in order to center the data we need to subtract
the mean of the vector points in the feature space:

$$mean = \frac{1}{n} \sum_{t=1}^{n} \phi(x_t) = \bar{\phi}$$

$$\Downarrow$$

$$v_i = \phi(x_i) - \bar{\phi} = \phi(x_i) - \frac{1}{n} \sum_{t=1}^{n} \phi(x_t)$$

and so we can see that the centered $k'_{i,j} = \langle v_i, v_j \rangle$ matrix
can be expressed as:

$$k'_{i,j} = \langle \phi(x_i) - \bar{\phi}, \phi(x_j) - \bar{\phi} \rangle$$

$$\Downarrow$$

$$k'_{i,j} = \langle \phi(x_i), \phi(x_j) \rangle - \langle \phi(x_i), \bar{\phi} \rangle - \langle \bar{\phi}, \phi(x_j) \rangle + \langle \bar{\phi}, \bar{\phi} \rangle$$

$$\Downarrow$$

$$k'_{i,j} = \langle \phi(x_i), \phi(x_j) \rangle - \frac{1}{n} \sum_{t=1}^{n} \langle \phi(x_i), \phi(x_t) \rangle - \frac{1}{n} \sum_{t=1}^{n} \langle \phi(x_t), \phi(x_j) \rangle$$
$$+ \sum_{t=1}^{n} \sum_{s=1}^{n} \langle \phi(x_t), \phi(x_s) \rangle$$

The original K matrix is $K_{i,j} \langle \phi(x_i), \phi(x_j) \rangle$ substituting
that in the expression above gives us the following expression:

$$k'_{i,j} = k_{i,j} - \frac{1}{n} \sum_{t=1}^{n} k_{i,j} - \frac{1}{n} \sum_{t=1}^{n} k_{t,j} + \frac{1}{n^2} \sum_{t=1}^{n} \sum_{s=1}^{n} k_{t,s}$$

in the above expression we can see that $k'_{i,j}$ is calculated
using $k_{i,j}$

2.

$$\frac{1}{n}\sum_{i=1}^{n}\phi(x_i)=0$$

$$S=\sum_{i=1}^{n}\phi(x_i)\phi(x_i)^T$$

The eigen vectors of S satisfy the following equasion:

$$S u_j = \lambda_j u_j$$

to prove that u_j is a linear combination of (x_i) we assume

$$u_j=\sum_{i=1}^{n} a_{j,i}\phi(x_i)$$

$$\Downarrow$$

$$S u_j = S\left(\sum_{i=1}^{n} a_{j,i}\phi(x_i)\right)=\lambda_j\left(\sum^{n} a_{j,i}\phi(x_i)\right)$$

$$=\sum_{i=1}^{n} a_{j,i}\sum_{t=1}^{n}\phi(x_t)\langle\phi(x_t),(x_i)\rangle$$

$$=\sum_{t=1}^{n}\phi(x_t)\sum_{i=1}^{n} a_{j,i}k_{t,i}$$

We know that $k_{t,i}=(\phi(x_t),\phi(x_i))$ so we use that to simplify the expression

$$\sum_{t=1}^{n}\phi(x_t)\left(\sum_{i=1}^{n} a_{j,i}\,k_{t,i}\right)=\sum_{t=1}^{n}\phi(x_t)\lambda_j a_{j,t}$$

if we match the coeffients we find that

$$\sum_{i=1}^{n} a_{j,i}k_{t,i}=\lambda_j a_{j,t}\implies k a_j=\lambda_j a_j$$

So we can see that a_j is il the vector of coefficents

3. Algorithm :

1. Calculate and center the kernel matrix K

2. Solve the following equasion $K a_j = \lambda_j a_j$ so that we find the eigenvalues and eigenvectors (use 2.)

3. Choose the top k eigenvectors that we found in the previous

4. normalize the chosen k eigenvalues and return them

4. $$z_j = \langle u_j, \varphi(x) \rangle = \left\langle \sum_{i=1}^{n} \alpha_{j,i} \varphi(x_i), \varphi(x) \right\rangle$$

$u_j = \sum_{i=1}^{n} \alpha_{j,i} \varphi(x_i)$

$$= \sum_{i=1}^{n} \alpha_{j,i} \langle \varphi(x_i), \varphi(x) \rangle$$

$$= \sum_{i=1}^{n} \underbrace{\alpha_{j,i}}_{\text{eigenvector coefficients}} \underbrace{k(x_i, x)}_{\text{kernel function}}$$

5. code in collab, yes it did better than PCA