Week 2 Graded assignment

Common data for Questions 1 and 2

A function k is defined as follows.

$$k : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$$
$$k(x_1, x_2) = x_1^T x_2$$

Question 1

Statement

Is k a valid kernel?

Options

(a)

Yes

(b)

No

Answer

(a)

Solution

Let ϕ be an identity transformation that is

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$$
$$\phi(x) = x$$

It is clear by the definition of given kernel that

$$k(x_1, x_2) = \phi(x_1)^T \phi(x_2)$$

It implies that k is a valid kernel.

Question 2

If k is the valid kernel, we apply it to the three-dimensional dataset to run the kernel PCA. Select the correct options.

Options

(a)

We cannot run the PCA as k is not a valid kernel.

(b)

It will be the same as PCA with no kernel.

(c)

It will be the same as the polynomial transformation of degree 2 and then run the PCA.

(d)

It will be the same as PCA with a third-degree polynomial kernel.

Answer

(b)

Solution

We have seen (in question 1) that k corresponds to the identity transformation. It implies that applying kernel and running PCA is same as standard PCA on the given dataset.

Question 3

Statement

Consider ten data points lying on a curve of degree two in a two-dimensional space. We run a kernel PCA with a polynomial kernel of degree two on the same data points. Choose the correct options.

Options

(a)

The transformed data points will lie on a 5-dimensional subspace of $\mathbb{R}^6$.

(b)

The transformed data points will lie on a 6-dimensional subspace of $\mathbb{R}^{10}$

(c)

There will be some $w \in \mathbb{R}^6$ that all of the data points are orthogonal to.

(d)

There will be some $w \in \mathbb{R}^{10}$ that all of the data points are orthogonal to.

Answer

(a), (c)

Solution

Since we are applying the polynomial kernel of degree two on the 2D dataset, the dataset will be transformed into a 6D feature space. (verify)

And the dataset is given to lying on a curve of degree two, the transformed dataset will live in the linear subspace of $\mathbb{R}^6$. and therefore, there will be some $w \in \mathbb{R}^6$ that all of the data points are orthogonal to.

Question 4

Statement

Which of the following matrices can not be appropriate matrix $K = X^T X$ for some data matrix X?

Options

(a)

$$\begin{bmatrix} 1 & 8 \\ 8 & -1 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 1 & 8 \\ 8 & 1 \end{bmatrix}$$

(c)

$$\begin{bmatrix} 1 & 8 \\ -8 & 1 \end{bmatrix}$$

(d)

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Answer

(a), (b) and (c)

Solution

We know that K matrix must be symmetric and positive semi-definite.

All the given matrices are symmetric.

We need to check whether they are positive semi-definite or not.

For that, we will check the eigenvalues of the matrices and if all the eigenvalues are non-negative, then the matrix is positive semi-definite.

Option (a)

$$A = \begin{bmatrix} 1 & 8 \\ 8 & -1 \end{bmatrix}$$

Let λ be the eigenvalue of A, then

$$|A - \lambda I| = 0$$
$$\begin{vmatrix} 1-\lambda & 8 \\ 8 & -1-\lambda \end{vmatrix} = 0$$
$$(1-\lambda)(-1-\lambda) = 64$$
$$\lambda^2 - 1 = 64$$
$$\lambda = \pm\sqrt{65}$$

Since A has a non-negative eigenvalue, A is not a positive semi-definite matrix.

Similarly, check for all the options.

Question 5

Statement

A function k is defined as

$$k : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$$
$$k(x_1, x_2) = (x_1^T x_2)^2$$

Is k a valid kernel?

Options

(a)

Yes

(b)

No

Answer

(a)

Solution

The given function is

$$k : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$$
$$k(x_1, x_2) = (x_1^T x_2)^2$$

Let $x_1 = [a_1, a_2]^T$ and $x_2 = [b_1, b_2]^T$ then

$$
\begin{aligned}
k(x_1, x_2) &= ([a_1, a_2][b_1, b_2]^T)^2 \\
&= (a_1 b_1 + a_2 b_2)^2 \\
&= a_1^2 b_1^2 + 2a_1 b_1 a_2 b_2 + a_2^2 b_2^2 \\
&= [a_1^2, \sqrt{2} a_1 a_2, a_2^2][b_1^2, \sqrt{2} b_1 b_2, b_2^2]^T \\
&= \phi(x_1)^T \phi(x_2)
\end{aligned}
$$

where $\phi : \mathbb{R}^2 \to \mathbb{R}$ such that $\phi([a_1, a_2]^T) = [a_1^2, \sqrt{2}a_1 a_2, a_2^2]^T$

It means there exists a transformation mapping ϕ such that $k(x_1, x_2) = \phi(x_1)^T \phi(x_2)$. Therefore, k is a valid kernel.

Question 6

Statement

Kernel PCA was run on the four data points $[1, 2]^T$, $[2, 3]^T$, $[2, -3]^T$, and $[4, 4]^T$ with the polynomial kernel of degree 2. What will be the shape of the matrix K? Notations are used as per lectures.

Options

(a)

2×2

(b)

4×4

(c)

6×6

(d)

None of the above

Answer

(b)

Solution

The K matrix is defined as $X^T X$ where X is a data matrix of shape (d, n). That is K matrix is of shape (n, n) where n is a number of examples.

It is given that $n = 4$. Therefore, shape of K matrix is $(4, 4)$

Question 7

Statement

Find the element at the index $(2,3)$ of the matrix K defined in Question 6. Take the points in the same order.

Options

(a)

-4

(b)

16

(c)

13

(d)

196

Answer

(b)

Solution

The polynomial kernel of degree 2 is given by

$$k(x_1, x_2) = (x_1^T x_2 + 1)^2$$

The $(2,3)th$ element of K matrix will be $k(x_2, x_3)$.

$$k(x_2, x_3) = ([2,3][2,-3]^T + 1)^2 = (-5+1)^2 = 16$$

Question 8

Statement

A dataset containing 200 examples in four-dimensional space has been transformed into higher dimensional space using the polynomial kernel of degree two. What will be the dimension of transformed feature space?

Answer

15 (No range required)

Solution

Let the features be x_1, x_2, x_3, and x_4. After the transformation of degree two, features will be $1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1^2, x_2^2, x_3^2$, and x_4^2. So, the dimension of transformed feature space will be 15.

Question 9

Statement

Let $x_1, x_2, \ldots, x_n$ be d-dimensional data points $(d > n)$ and X be the matrix of shape $d \times n$ containing the data points. The k^{th} largest eigenvalue and corresponding unit eigenvector of $X^T X$ is λ and α_k, respectively. What will be the projection of x_i on the k^{th} principal component?

Options

(a)

$x_i^T \alpha_k$

(b)

$\frac{x_i^T \alpha_k}{\lambda}$

(c)

$\frac{x_i^T X \alpha_k}{\sqrt{\lambda}}$

(d)

$\frac{x_i^T X \alpha_k}{\sqrt{n\lambda}}$

Answer

(c)

Solution

If the k^{th} largest eigenvalue and corresponding unit eigenvector of $X^T X$ is λ and α_k, respectively, the k^{th} principal component will be $\frac{X\alpha_k}{\sqrt{\lambda}}$.

Therefore, the projection of the point x_i on the k^{th} principal component will be $\frac{x_i^T X \alpha_k}{\sqrt{\lambda}}$.

Question 10

Statement

Let k_1 and k_2 be two valid kernels. Is $3k_1 + 5k_2$ a valid kernel?

Options

(a)

Yes

(b)

No

Answer

(a)

Solution

Let k_1 and k_2 be two valid kernels defined on $\mathbb{R}^d \times \mathbb{R}^d$, it implies that they satisfies the following two properties

(1) k_1 and k_2 are symmetric functions that is

$$k_1(x_1, x_2) = k_1(x_2, x_1) \quad \forall x_1, x_2 \in \mathbb{R}^d \quad \text{and}$$
$$k_2(x_1, x_2) = k_2(x_2, x_1) \quad \forall x_1, x_2 \in \mathbb{R}^d$$

(2) For any $x \in \mathbb{R}^n$, we have

$$x^T K_1 x \geq 0 \quad \text{and}$$
$$x^T K_2 x \geq 0$$

where,

$K_1 = [k_1(x_i, x_j)]_{n \times n}$ and $K_2 = [k_2(x_i, x_j)]_{n \times n}$ are K matrices corresponding to kernels k_1 and k_2, respectively.

Assume that $k = 3k_1 + 5k_2$

To show: $k(x_1, x_2) = k(x_2, x_1) \quad \forall x_1, x_2 \in \mathbb{R}^d$

and

$x^T K x \geq 0 \quad \forall x \in \mathbb{R}^n$, where $K = [k(x_i, x_j)]_{n \times n}$ is K matrices corresponding to kernel k

Now,

$$\begin{aligned} k(x_1, x_2) &= 3k_1(x_1, x_2) + 5k_2(x_1, x_2) \\ &= 3k_1(x_2, x_1) + 5k_2(x_2, x_1) \\ &= k(x_2, x_1) \end{aligned}$$

and

$$\begin{aligned} x^T K x &= x^T (3K_1 + 5K_2) x \\ &= 3x^T K_1 x + 5x^T K_2 x \geq 0 \end{aligned}$$

It implies that $3k_1 + 5k_2$ is a valid kernel.