Practice assignment

Question: 1

Statement

Assume that $w_k;\ k = 1, 2, \ldots d$ are d principal components corresponding to nonzero eigenvalues of the D-dimensional centered data points $x_i;\ i = 1, 2, \ldots, n$.

Statement 1: each x_i can be written as a linear combination of $w_k s$.

Statement 2: each w_k can be written as a linear combination of $x_i s$.

Options

(a)

Statement 1 is correct but statement 2 is incorrect.

(b)

Statement 1 is incorrect but statement 2 is correct.

(c)

Both statements are correct.

(d)

Both statements are incorrect.

Answer:

(c)

Solution

In the first week, we have seen that residues after d iterations become zero that is

$$\begin{aligned} & x_i - (x_i^T w_1 + x_i^T w_2 + \cdots + x_i^T w_d) = 0 \\ \Rightarrow \quad & x_i = x_i^T w_1 + x_i^T w_2 + \cdots + x_i^T w_d \end{aligned}$$

it implies that each x_i can be written as a linear combination of $w_k s$.

We know that the eigenvectors of the covariance matrix C are the principal components of the dataset and by the definition of eigenvectors, we have

$$\begin{aligned} Cw_k &= \lambda_k w_k \\ \Rightarrow \quad \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right) w_k &= \lambda_k w_k \\ \Rightarrow \quad w_k &= \sum_{i=1}^{n}\left(\frac{x_i^T w_k}{n\lambda_k}\right) x_i \end{aligned}$$

That is ach w_k can be written as a linear combination of $x_i s$.

Question: 2

Statement

A transformation mapping ϕ is defined as

$$\phi : \mathbb{R} \to \mathbb{R}^4$$
$$\phi(x) = [x^3, \sqrt{3}x^2, \sqrt{3}x, 1]^T$$

Which of the following options are the same as $\phi(x_1)^T\phi(x_2)$ for two points $x_1, x_2 \in \mathbb{R}$?

Hint: Rather than doing the calculation, try to figure out the appropriate kernel function.

(a)

$(x_1 x_2 + 1)^3$

(b)

$(x_2 x_1 + 1)^3$

(c)

$(x_1 x_2 + 1)^4$

(d)

$\phi(x_2)^T\phi(x_1)$

Answer:

(a), (b), (d)

Solution

It is easy to verify that

$$\phi(x_1)^T\phi(x_2) = x_1^3 x_2^3 + 3x_1^2 x_2^2 + 3x_1 x_2 + 1 = (x_1 x_2 + 1)^3$$

It shows that the polynomial kernel of degree three refers to the given transformation ϕ. And since the dot product is commutative, we can check that options (a), (b), and (d) refer to the same expression.

Therefore the correct answers are options (a), (b), and (d).

Question: 3

Statement

Let C be the covariance matrix of n data points in d-dimensional space. Assume that the data points are mean-centered. If $2, 5,$ and 7 are the only non-zero eigenvalues of C, what will be the non-zero eigenvalues of X^TX, where X is the matrix of shape (d, n) containing the data points?

Options

(a)

$2, 5, 7$

(b)

$2d, 5d, 7d$

(c)

$2n, 5n, 7n$

(d)

Can not be determined

Answer: C

Solution

The covariance matrix is defined as $\frac{1}{n}XX^T$ and the nonzero eigenvalues of $\frac{1}{n}XX^T$ are given to be $2, 5$, and 7.

$\Rightarrow$ nonzero eigenvalues of XX^T will be $2n, 5n$ and $7n$.

Since all the nonzero eigenvalues of XX^T and X^TX are the same, the nonzero eigenvalues of X^TX are $2n, 5n$, and $7n$.

Common data for Questions 4 and 5

Statement

Consider an image dataset matrix X of shape (d, n) with $d > n$. The kth principal component of the dataset can be written as $w_k = \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n$, where, the vector x_i is the ith data point. The k^{th} largest eigenvalue and the corresponding eigenvector of X^TX are 4 and $[\frac{1}{\sqrt{51}}, \frac{3}{\sqrt{51}}, \frac{4}{\sqrt{51}}, \frac{5}{\sqrt{51}}]^T$, respectively.

Question 4

Statement

What will be the value of α_1?

Options

(a)

4

(b)

$$\frac{1}{\sqrt{51}}$$

(c)

$$\frac{1}{2\sqrt{51}}$$

(d)

$[\frac{1}{\sqrt{51}}, \frac{3}{\sqrt{51}}, \frac{4}{\sqrt{51}}, \frac{5}{\sqrt{51}}]^T$

(e)

$[\frac{1}{2\sqrt{51}}, \frac{3}{2\sqrt{51}}, \frac{4}{2\sqrt{51}}, \frac{5}{2\sqrt{51}}]^T$

Answer:

C

Solution

We know that the k^{th} principal component can be written as a linear combination of the data points that is

$$w_k = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n$$

And the vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n]$ can be obtained by eigen decomposition of $X^T X$ as follows:

If the k^{th} largest eigenvalue and the corresponding unit eigenvector of $X^T X$ are λ_k and β_k, respectively then

$$\alpha = \frac{\beta_k}{\sqrt{\lambda_k}}$$

Therefore, by the given information, we can say that

$$[\alpha_1, \alpha_2, \ldots, \alpha_n]^T = \frac{1}{\sqrt{4}}[\frac{1}{\sqrt{51}}, \frac{3}{\sqrt{51}}, \frac{4}{\sqrt{51}}, \frac{5}{\sqrt{51}}]^T$$
$$\Rightarrow \alpha_1 = \frac{1}{2\sqrt{51}}$$

Question 5

Statement

What will be the k^{th} largest eigenvalue of the covariance matrix $\frac{1}{4}XX^T$?

Note that $n = 4$ as the length of the eigenvector of X^TX is 4.

Answer:

1

Solution

The k^{th} largest eigenvalue of $X^TX = 4$

The nonzero eigenvalues of X^TX and XX^T are the same.

$\Rightarrow$ The k^{th} largest eigenvalue of $XX^T = 4$

$\Rightarrow$ The k^{th} largest eigenvalue of $\frac{1}{4}XX^T = 1$

Question: 6

Statement

A function k is defined as

$$k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$$
$$k([x_1, x_2]^T, [y_1, y_2]^T) = x_1^2y_1^2 + x_2^2y_2^2$$

Is k a valid kernel?

Hint: Try to find out the appropriate ϕ.

Options

(a)

Yes

(b)

No

Answer:

A

Solution

If we can find an appropriate transformation mapping ϕ such that

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

Then we can conclude that k is a valid kernel.

The given kernel is

$$
\begin{aligned}
k : \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R} \\
k([x_1, x_2]^T, [y_1, y_2]^T) \qquad &= x_1^2 y_1^2 + x_2^2 y_2^2 \\
&= [x_1^2, x_2^2]^T [y_1^2, y_2^2]
\end{aligned}
$$

If we define a function $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$\phi([x_1, x_2]) = [x_1^2, x_2^2]$$

then we can say that

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

It implies that k is a valid kernel.

Question: 7

Statement

A dataset of 1000 second-hand cars has four features: kilometers driven (x_1), mileage (x_2), the present price of the car (x_3), and the selling price (x_4). The selling price seems to have the following relationship (approximate) with the other three features.

$$x_4 = x_1^2 x_3 + 2x_2$$

If we want to project the dataset into a lower dimensional space, which of the following task would be most appropriate?

Options

(a)

Standard PCA

(b)

Kernel PCA with a polynomial kernel of degree 2

(c)

Kernel PCA with a polynomial kernel of degree 3

(d)

Kernel PCA with a polynomial kernel of degree 4

Answer

(c)

Solution

Notice that the features are not linearly related. The feature x_4 is related to other features and the relationship is cubic in nature.

That is why if we transform the dataset into a higher dimension using the degree three polynomial, then the dataset may live in a linear subspace.

Therefore, kernel PCA with a polynomial kernel of degree 3 will be an appropriate task.

Question 8

Statement

Abhishek runs a kernel PCA on a dataset containing n examples with d features. Which of the following strategy he should follow to center the data points?

strategy 1: First center the dataset using the mean and then apply the kernel.

Strategy 2: First apply the kernel and then center the matrix.

Options

(a)

Strategy 1

(b)

Strategy 2

(c)

Both strategies are the same

Answer

(b)

Solution

Applying transformation on the centered dataset is not mandatory to give the centered transformed dataset. For example, consider a centered dataset containing four points in two dimensions.

$$X = \begin{bmatrix} 1 & -1 & 3 & -3 \\ 2 & -2 & 4 & -4 \end{bmatrix}$$

Now apply the transformation $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$\phi([x_1, x_2]) = [x_1^2, x_2^2]$$

The transformed dataset will be

$$X' = \begin{bmatrix} 1 & 1 & 9 & 9 \\ 4 & 4 & 16 & 16 \end{bmatrix}$$

Clearly, this dataset is not centered. Therefore, strategy 2 is best suited.

Question 9

Statement

A dataset containing 1000 points in 3-dimensional space is run through the kernel PCA with the polynomial kernel of degree p. If the transformed dataset lives in a ten-dimensional space, what will be the value of p?

Answer

2 (No range required)

Solution

Let x_1, x_2 and x_3 are three features. If we use $p = 2$ the features will be

$1, x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$ that is the transformed space will be ten-dimensional.

Question 10

Statement

A dataset containing 1000 examples in 10-dimensional space is projected into other dimension space using kernel PCA with the following kernel.

$$k(x_1, x_2) = \exp\left(\frac{-||x_1 - x_2||^2}{4}\right)$$

What will be the dimension of the projected dataset?

Options

(a)

10

(b)

40

(c)

∞

(d)

Can not be determined

Answer

(c)

Solution

The given kernel is the gaussian kernel, which will lead to infinite dimension.