MLT: Week 2

Kernel PCA

Sherry Thomas

Kernel PCA Algorithm

Let's take a dataset X where

d: no. of features

n: no. of datapoints

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

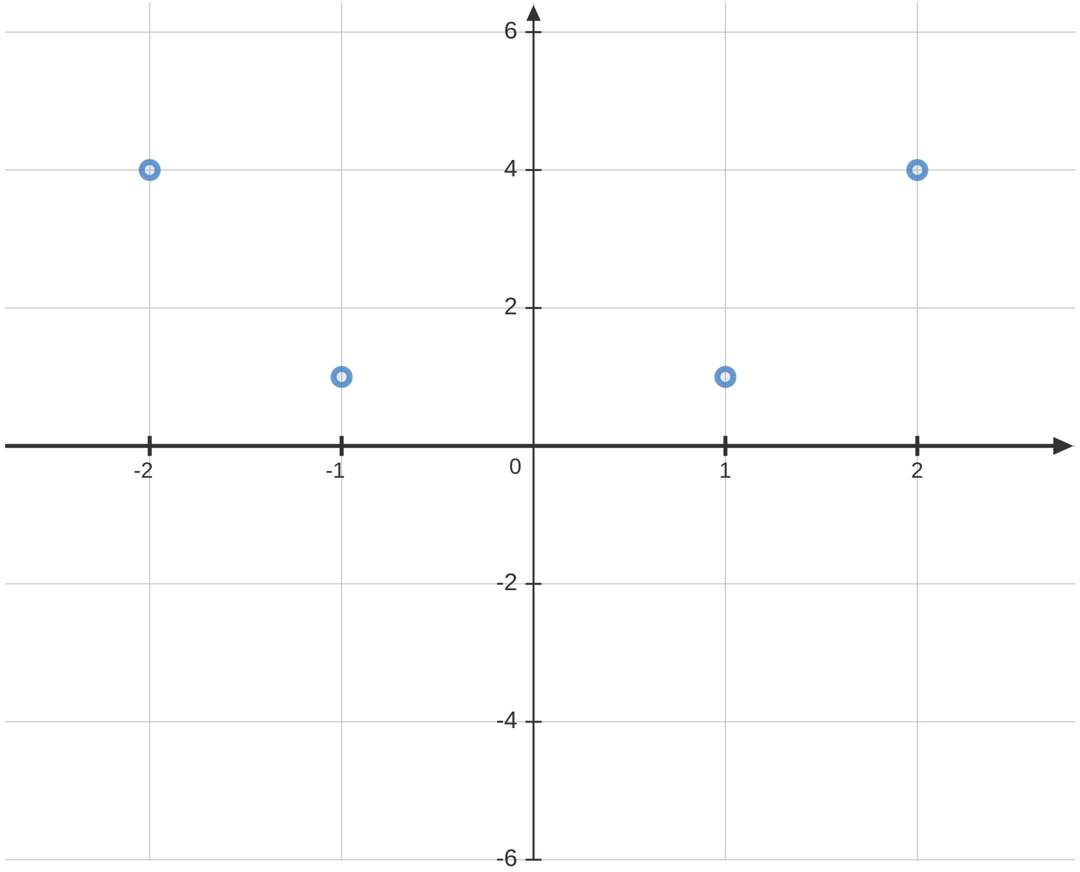
Kernel PCA Algorithm

Let's take a dataset X where

d: no. of features

n: no. of datapoints

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



Step 1: Calculate $\mathrm{K} \in \mathbb{R}^{n \times n}$ using a kernel function where $\mathrm{K}_{ij} = k(x_i, x_j)$.

Step 1: Calculate $\mathrm{K} \in \mathbb{R}^{n \times n}$ using a kernel function where $\mathrm{K}_{ij} = k(x_i, x_j)$.

Given the apparent quadratic nature of the dataset, we opt for a polynomial kernel with a degree of two.

Step 1: Calculate $\mathrm{K} \in \mathbb{R}^{n \times n}$ using a kernel function where $\mathrm{K}_{ij} = k(x_i, x_j)$.

Given the apparent quadratic nature of the dataset, we opt for a polynomial kernel with a degree of two.

$$k(\mathrm{x}_i, \mathrm{x}_j) = \left(\mathrm{x}_i^T \mathrm{x}_j \ + \ 1\right)^2$$

Step 1: Calculate $\mathrm{K} \in \mathbb{R}^{n \times n}$ using a kernel function where $\mathrm{K}_{ij} = k(x_i, x_j)$.

Given the apparent quadratic nature of the dataset, we opt for a polynomial kernel with a degree of two.

$$k(\mathrm{x}_i, \mathrm{x}_j) = \left(\mathrm{x}_i^T \mathrm{x}_j \ + \ 1\right)^2$$

$$k(\mathrm{x}_1, \mathrm{x}_1) = \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \right)^2 = (\ 2 + 1\ )^2 = 9$$

Step 1: Calculate $\mathrm{K} \in \mathbb{R}^{n \times n}$ using a kernel function where $\mathrm{K}_{ij} = k(x_i, x_j)$.

Given the apparent quadratic nature of the dataset, we opt for a polynomial kernel with a degree of two.

$$k(\mathrm{x}_i, \mathrm{x}_j) = \left(\mathrm{x}_i^T \mathrm{x}_j \ + \ 1\right)^2$$

$$k(\mathrm{x}_1, \mathrm{x}_1) = \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \right)^2 = (\ 2 + 1\ )^2 = 9$$

$$k(\mathrm{x}_1, \mathrm{x}_2) = \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 1 \right)^2 = (\ 6 + 1\ )^2 = 49$$

Step 1: Calculate $\mathrm{K} \in \mathbb{R}^{n \times n}$ using a kernel function where $\mathrm{K}_{ij} = k(x_i, x_j)$.

Given the apparent quadratic nature of the dataset, we opt for a polynomial kernel with a degree of two.

$$k(\mathrm{x}_i, \mathrm{x}_j) = \left(\mathrm{x}_i^T \mathrm{x}_j \; + \; 1\right)^2$$

$$k(\mathrm{x}_1, \mathrm{x}_1) = \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \right)^2 = (\ 2 + 1\ )^2 = 9$$

$$k(\mathrm{x}_1, \mathrm{x}_2) = \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 1 \right)^2 = (\ 6 + 1\ )^2 = 49$$

$$\therefore\ \mathrm{K} = \begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix}$$

Step 2: Center the kernel using the following formula.

$$\mathrm{K}^C = \mathrm{K} - \mathrm{IK} - \mathrm{KI} + \mathrm{IKI}$$

where K^C is the centered kernel, and $\mathrm{I} \in \mathbb{R}^{n \times n}$ where all the elements are $\frac{1}{n}$.

Step 2: Center the kernel using the following formula.

$$\mathrm{K}^C = \mathrm{K} - \mathrm{IK} - \mathrm{KI} + \mathrm{IKI}$$

where K^C is the centered kernel, and $\mathrm{I} \in \mathbb{R}^{n\times n}$ where all the elements are $\frac{1}{n}$.

$$\begin{aligned}
\mathrm{K}^C = &\begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix} - \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}\begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix} \\
&- \begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix}\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} \\
&+ \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}\begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix}\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}
\end{aligned}$$

Step 2: Center the kernel using the following formula.

$$\mathrm{K}^C = \mathrm{K} - \mathrm{IK} - \mathrm{KI} + \mathrm{IKI}$$

where K^C is the centered kernel, and $\mathrm{I} \in \mathbb{R}^{n \times n}$ where all the elements are $\frac{1}{n}$.

$$\begin{aligned}\mathrm{K}^C = &\begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix} - \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}\begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix} \\ &- \begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix}\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} \\ &+ \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}\begin{bmatrix} 9 & 49 & 1 & 9 \\ 49 & 441 & 9 & 169 \\ 1 & 9 & 9 & 49 \\ 9 & 169 & 49 & 441 \end{bmatrix}\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}\end{aligned}$$

$$\mathrm{K}^C = \begin{bmatrix} 67 & -43 & 59 & -83 \\ -43 & 199 & -83 & -73 \\ 59 & -83 & 67 & -43 \\ -83 & -73 & -43 & 199 \end{bmatrix}$$

Step 3: Compute the eigenvectors $\{\beta_1, \beta_2, \dots, \beta_n\}$ and eigenvalues $\{n\lambda_1, n\lambda_2, \dots, n\lambda_n\}$ of K^C and normalize to get

$$\forall u \quad \alpha_u = \frac{\beta_{\mathrm{u}}}{\sqrt{n\lambda_u}}$$

Step 3: Compute the eigenvectors $\{\beta_1, \beta_2, \dots, \beta_n\}$ and eigenvalues $\{n\lambda_1, n\lambda_2, \dots, n\lambda_n\}$ of K^C and normalize to get

$$\forall u \quad \alpha_u = \frac{\beta_{\mathrm{u}}}{\sqrt{n\lambda_u}}$$

$$(n\lambda,\ \beta) = \ Solver\left(\mathrm{K}^C\right)$$

Step 3: Compute the eigenvectors $\{\beta_1, \beta_2, \ldots, \beta_n\}$ and eigenvalues $\{n\lambda_1, n\lambda_2, \ldots, n\lambda_n\}$ of K^C and normalize to get

$$\forall u \quad \alpha_u = \frac{\beta_{\mathrm{u}}}{\sqrt{n\lambda_u}}$$

$$(n\lambda,\ \beta) = \ Solver\left(\mathrm{K}^C\right)$$

$$n\lambda = \left[\, 277.927 \ \ 252 \ \ 2.072 \ \ 0 \,\right]$$

Step 3: Compute the eigenvectors $\{\beta_1, \beta_2, \dots, \beta_n\}$ and eigenvalues $\{n\lambda_1, n\lambda_2, \dots, n\lambda_n\}$ of K^C and normalize to get

$$\forall u \quad \alpha_u = \frac{\beta_{\mathrm{u}}}{\sqrt{n\lambda_u}}$$

$$(n\lambda,\ \beta) = \ Solver\left(\mathrm{K}^C\right)$$

$$n\lambda = \begin{bmatrix} 277.927 & 252 & 2.072 & 0 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 0.10365278 & -0.5 & -0.69946844 & 0.5 \\ 0.69946844 & 0.5 & 0.10365278 & 0.5 \\ -0.10365278 & -0.5 & 0.69946844 & 0.5 \\ -0.69946844 & 0.5 & -0.10365278 & 0.5 \end{bmatrix}$$

$$\alpha_1 = \frac{\beta_1}{\sqrt{n\lambda_1}} = \begin{bmatrix} 0.10365278 \\ 0.69946844 \\ -0.10365278 \\ -0.69946844 \end{bmatrix} / \sqrt{277.927} = \begin{bmatrix} 0.00621749 \\ 0.0419568 \\ -0.00621749 \\ -0.0419568 \end{bmatrix}$$

$$\alpha_1 = \frac{\beta_1}{\sqrt{n\lambda_1}} = \begin{bmatrix} 0.10365278 \\ 0.69946844 \\ -0.10365278 \\ -0.69946844 \end{bmatrix} / \sqrt{277.927} = \begin{bmatrix} 0.00621749 \\ 0.0419568 \\ -0.00621749 \\ -0.0419568 \end{bmatrix}$$

$$\alpha_2 = \frac{\beta_2}{\sqrt{n\lambda_2}} = \begin{bmatrix} -0.5 \\ 0.5 \\ -0.5 \\ 0.5 \end{bmatrix} / \sqrt{252} = \begin{bmatrix} -0.03149704 \\ 0.03149704 \\ -0.03149704 \\ 0.03149704 \end{bmatrix}$$

$$\alpha_1 = \frac{\beta_1}{\sqrt{n\lambda_1}} = \begin{bmatrix} 0.10365278 \\ 0.69946844 \\ -0.10365278 \\ -0.69946844 \end{bmatrix} / \sqrt{277.927} = \begin{bmatrix} 0.00621749 \\ 0.0419568 \\ -0.00621749 \\ -0.0419568 \end{bmatrix}$$

$$\alpha_2 = \frac{\beta_2}{\sqrt{n\lambda_2}} = \begin{bmatrix} -0.5 \\ 0.5 \\ -0.5 \\ 0.5 \end{bmatrix} / \sqrt{252} = \begin{bmatrix} -0.03149704 \\ 0.03149704 \\ -0.03149704 \\ 0.03149704 \end{bmatrix}$$

$$\alpha_3 = \frac{\beta_3}{\sqrt{n\lambda_3}} = \begin{bmatrix} -0.69946844 \\ 0.10365278 \\ 0.69946844 \\ -0.10365278 \end{bmatrix} / \sqrt{2.072} = \begin{bmatrix} -0.48587288 \\ 0.0720005 \\ -0.48587288 \\ -0.0720005 \end{bmatrix}$$

Step 3: Compute $\sum_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \quad \forall k$

$$\phi(\mathrm{x}_i)^T \mathrm{w} \in \mathbb{R}^k \rightarrow \left[\begin{array}{cccc} \sum_{j=1}^{n} \alpha_{1j} \mathrm{K}_{ij}^{C} & \sum_{j=1}^{n} \alpha_{2j} \mathrm{K}_{ij}^{C} & \ldots & \sum_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \end{array} \right]$$

Step 3: Compute $\sum_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \quad \forall k$

$$\phi(\mathrm{x}_i)^T \mathrm{w} \in \mathbb{R}^k \rightarrow \left[\sum_{j=1}^{n} \alpha_{1j} \mathrm{K}_{ij}^{C} \quad \sum_{j=1}^{n} \alpha_{2j} \mathrm{K}_{ij}^{C} \quad \dots \quad \sum_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \right]$$

$$\phi(\mathrm{X})^T \mathrm{w} \in \mathbb{R}^{n \times k} \rightarrow \ \mathrm{K} \ \alpha$$

Step 3: Compute $\sum_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \quad \forall k$

$$\phi(\mathrm{x}_i)^T \mathrm{w} \in \mathbb{R}^k \rightarrow \begin{bmatrix} \sum_{j=1}^{n} \alpha_{1j} \mathrm{K}_{ij}^{C} & \sum_{j=1}^{n} \alpha_{2j} \mathrm{K}_{ij}^{C} & \dots & \sum_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \end{bmatrix}$$

$$\phi(\mathrm{X})^T \mathrm{w} \in \mathbb{R}^{n \times k} \rightarrow \ \mathrm{K} \ \alpha$$

$$\phi(\mathrm{X})^T \mathrm{w} \in \mathbb{R}^{n \times k} \rightarrow \ \mathrm{K} \ \alpha \rightarrow \begin{bmatrix} 67 & -43 & 59 & -83 \\ -43 & 199 & -83 & -73 \\ 59 & -83 & 67 & -43 \\ -83 & -73 & -43 & 199 \end{bmatrix} \begin{bmatrix} 0.00621749 & -0.03149704 & -0.48587288 \\ 0.0419568 & 0.03149704 & 0.0720005 \\ -0.00621749 & -0.03149704 & -0.48587288 \\ -0.0419568 & 0.03149704 & -0.0720005 \end{bmatrix}$$

Step 3: Compute $\sum_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \;\; \forall k$

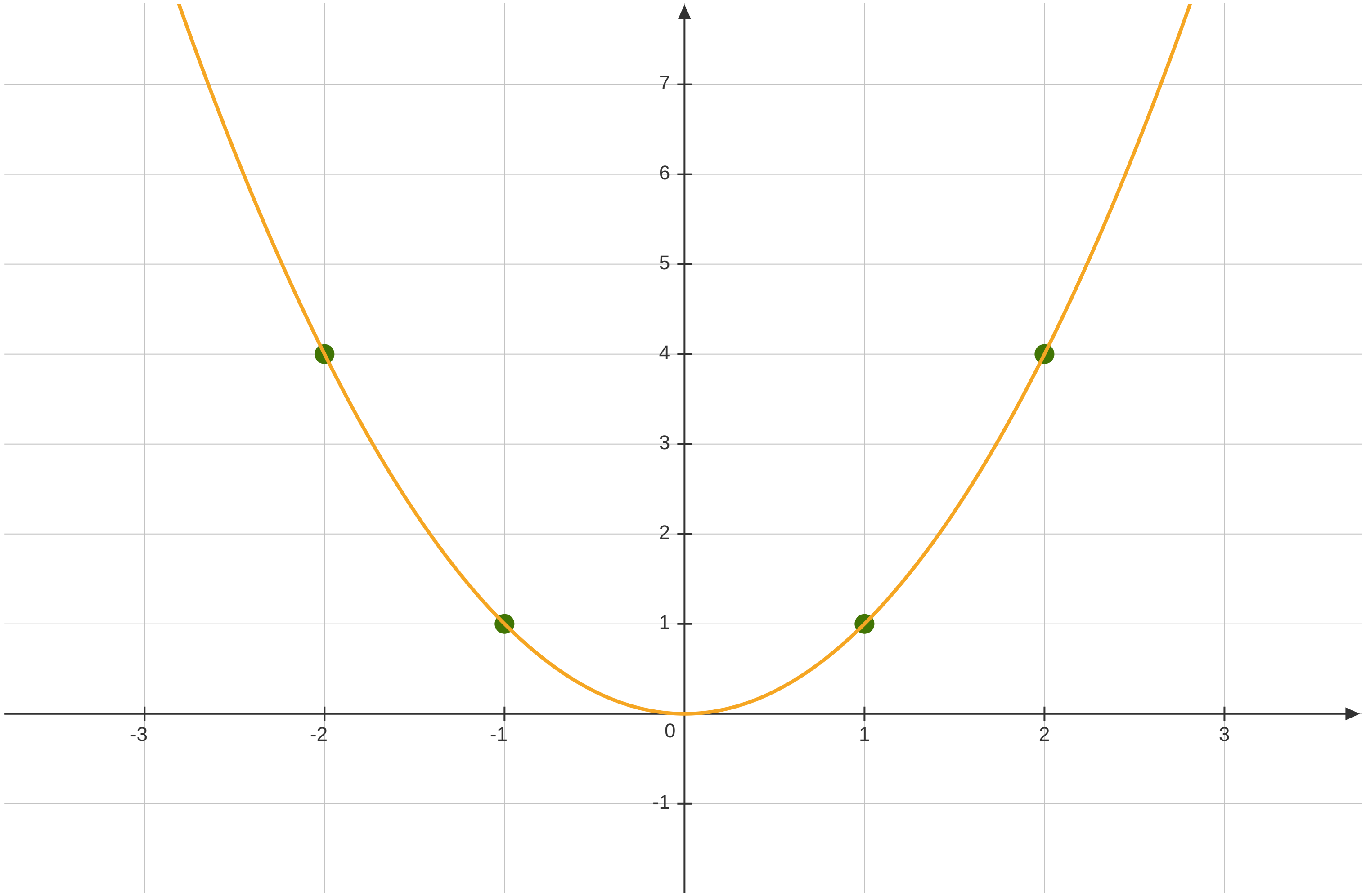
$$\phi(\mathrm{x}_i)^T \mathrm{w} \in \mathbb{R}^k \rightarrow \begin{bmatrix} \sum\limits_{j=1}^{n} \alpha_{1j} \mathrm{K}_{ij}^{C} & \sum\limits_{j=1}^{n} \alpha_{2j} \mathrm{K}_{ij}^{C} & \dots & \sum\limits_{j=1}^{n} \alpha_{kj} \mathrm{K}_{ij}^{C} \end{bmatrix}$$

$$\phi(\mathrm{X})^T \mathrm{w} \in \mathbb{R}^{n \times k} \rightarrow \ \mathrm{K}\ \alpha$$

$$\phi(\mathrm{X})^T \mathrm{w} \in \mathbb{R}^{n \times k} \rightarrow \ \mathrm{K}\ \alpha \rightarrow \begin{bmatrix} 67 & -43 & 59 & -83 \\ -43 & 199 & -83 & -73 \\ 59 & -83 & 67 & -43 \\ -83 & -73 & -43 & 199 \end{bmatrix} \begin{bmatrix} 0.00621749 & -0.03149704 & -0.48587288 \\ 0.0419568 & 0.03149704 & 0.0720005 \\ -0.00621749 & -0.03149704 & -0.48587288 \\ -0.0419568 & 0.03149704 & -0.0720005 \end{bmatrix}$$

$$\phi(\mathrm{X})^T \mathrm{w} \in \mathbb{R}^{n \times k} \rightarrow \ \mathrm{K}\ \alpha \rightarrow \begin{bmatrix} 1.72801191 & -7.93725393 & -1.00696319 \\ 11.66094908 & 7.93725393 & 0.14921979 \\ -1.72801191 & -7.93725393 & 1.00696319 \\ -11.66094908 & 7.93725393 & -0.14921979 \end{bmatrix}$$

Data Representation using Kernel PCA



Kernel Functions

A function that maps $k\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, and is "valid", is called a **Kernel Function**.

Kernel Functions

A function that maps $k\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, and is "valid", is called a **Kernel Function**.

Proof of a "Valid" Kernel:

- **Method 1**: Exhibit the map to ϕ explicitly. [may be hard]
- **Method 2**: Using Mercer's Theorem:

  $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a "valid" kernel if and only if:

  - k is symmetric i.e $k(x, x') = k(x', x)$
  - For any dataset $\{x_1, x_2, ..., x_n\}$, the matrix $K \in \mathbb{R}^{n \times n}$, where $\mathrm{K}_{ij} = k(i, j)$, is Positive Semi-Definite.

Kernel Functions

A function that maps $k \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, and is "valid", is called a **Kernel Function**.

Proof of a "Valid" Kernel:

- **Method 1**: Exhibit the map to ϕ explicitly. [may be hard]
- **Method 2**: Using Mercer's Theorem:

  $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a "valid" kernel if and only if:

  - k is symmetric i.e $k(x, x') = k(x', x)$
  - For any dataset $\{x_1, x_2, \ldots, x_n\}$, the matrix $K \in \mathbb{R}^{n \times n}$, where $\mathrm{K}_{ij} = k(i, j)$, is Positive Semi-Definite.

Two Popular Kernel Functions:

- **Polynomial Kernel**: $k(x, x') = (x^T x' + 1)^p$
- **Radial Basis Function Kernel** or **Gaussian Kernel**: $k(x, x') = exp(-\frac{||x - x'||^2}{2\sigma^2})$

Question 1:

Q. Consider 1000 data points belonging to a d-dimensional space have a non-linear relationship. We apply Kernel PCA to reduce the dimension of data points and take first k principal components. Can the value of k be larger than d?

Question 1:

Q. Consider 1000 data points belonging to a d-dimensional space have a non-linear relationship. We apply Kernel PCA to reduce the dimension of data points and take first k principal components. Can the value of k be larger than d?

A. Let the 1000 data points belong to a 2-dimensional space (x, y).
On applying a polynomial kernel of degree two, we get

$$\phi\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = \left(\begin{bmatrix} x_1 & y_1 \end{bmatrix}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + 1\right)^2 = (x_1x_2 + y_1y_2 \ + 1)^2 = x_1^2x_2^2 + 2x_1y_1x_2y_2 + 2x_1x_2 + y_1^2y_2^2 + 2y_1y_2 + 1$$

The original 2-dimensional space undergoes a mapping to a 6-dimensional space.
Applying PCA on this transformed dataset yields 6 principal components.
Hence, we **may** choose a value of k that exceeds d.

Question 2:

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

Question 2:

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

A. In this question, $d = 3,\ p = 2$. To find the dimension of the transformed feature space, we use the following formula:

$$^{d+p}C_d = {}^{3+2}C_2 = \frac{5!}{3!2!} = \frac{5 \times 4}{2} = 10$$

Therefore, when we map a three-dimensional space to a higher-dimensional space using a polynomial kernel of degree two, we get a transformed feature space of dimension ten.

Question 3:

Q. A kernel k is defined as:

$$k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

$$k\Big([x_1, x_2]^T, [y_1, y_2]^T\Big) = 1 + 2x_1^2 y_1^2 + 2x_2^2 y_2^2$$

Which of the following transformation mappings correspond to this kernel function?

a. $\phi([x_1, x_2]) = \begin{bmatrix} 1 & x_1^2 & x_2^2 \end{bmatrix}^T$

b. $\phi([x_1, x_2]) = \begin{bmatrix} 1 & \sqrt{2}x_1^2 + \sqrt{2}x_2^2 \end{bmatrix}^T$

c. $\phi([x_1, x_2]) = \begin{bmatrix} 1 & \sqrt{2}x_1^2 & \sqrt{2}x_2^2 \end{bmatrix}^T$

d. $\phi([x_1, x_2]) = \begin{bmatrix} 1 & x_1^2 & x_2^2 \end{bmatrix}^T$

A. Using the transformation in option (c), we get

$$k\Big([x_1, x_2]^T, [y_1, y_2]^T\Big) = \phi([x_1, x_2])^T \phi([y_1, y_2])$$

$$= \begin{bmatrix} 1 & \sqrt{2}x_1^2 & \sqrt{2}x_2^2 \end{bmatrix} \begin{bmatrix} 1 \\ \sqrt{2}y_1^2 \\ \sqrt{2}y_2^2 \end{bmatrix} = \begin{bmatrix} 1 + \left(\sqrt{2}x_1^2\right)\left(\sqrt{2}y_1^2\right) + \left(\sqrt{2}x_2^2\right)\left(\sqrt{2}y_2^2\right) \end{bmatrix}$$

$$= \begin{bmatrix} 1 + 2x_1^2y_1^2 + 2x_2^2y_2^2 \end{bmatrix}$$