Set M2, July 16, 2023

Consider the following kernel function:

$$k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$$
$$k([x_1, x_2]^T, [y_1, y_2]^T) = 1 + x_1y_1 + x_2y_2 + x_1^2y_1^2x_2y_2 + x_1y_1x_2^2y_2^2$$

Find the appropriate transformation mapping ϕ for the given kernel.

$\phi(X^T) = [1, x_1, x_2, x_1^2 x_2, x_2^2 x_1]$ $\quad (X = [x_1, x_2])$

Consider the following Hard K-means clustering problem with four points: $P(1,1), Q(2,1), R(4,3)$, and $S(5,4)$. Consider the number of clusters to be $k=2$ and the initial centroids to be $C_1 = (0,0)$ and $C_2 = (4,4)$. After how many iterations will the algorithm terminate and what will be the final centroids? use euclidean distance metric to calculate distance.

P and Q will be assigned to C_1. (Cuz they closest to C_1)
R and S will be assigned to C_2 (" " " " C_2)

On first Iteration: All points are assigned | On second iteration: New means are calculated, points are reassigned

[In our case points remain in the same cluster, so algo ends here]

$\mu_1 = \frac{\begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}2\\1\end{bmatrix}}{2} = \begin{bmatrix}1.5\\1\end{bmatrix}$ | $\mu_2 = \frac{\begin{bmatrix}4\\3\end{bmatrix} + \begin{bmatrix}5\\4\end{bmatrix}}{2} = \begin{bmatrix}4.5\\3.5\end{bmatrix}$ | Iterations = 2

Select all true statements regarding standard PCA applied on a centered dataset.

Options :

6406531929899. ✔ The variance of the dataset along the first principal component is maximum. Fact

6406531929900. ✖ The variance of the dataset along the first principal component is minimum.

6406531929901. ✔ The first principal component is the unit norm eigenvector corresponding to the largest eigenvalue of the covariance matrix. Fact

6406531929902. ✖ The first principal component is the unit norm eigenvector corresponding to the smallest eigenvalue of the covariance matrix.

Consider a dataset of 2000 points all of which lie in $\mathbb{R}^{30}$. If all these points lie in four dimensional low subspace then after applying the PCA algorithm which of the following statements is/are true?

Options :

6406531929903. ✖ The covariance matrix corresponding to this dataset has 30 non-zero eigenvalues.

6406531929904. ✔ The covariance matrix corresponding to this dataset has only four non-zero eigenvalues.

6406531929905. ✔ The residues will become zero after four rounds

6406531929906. ✔ By utilizing four principal component vectors and their corresponding coefficients, we have the ability to reconstruct the complete dataset using representatives.

Which of the following statements about kernel PCA is/are always true?

Options :

6406531929907. ✖ The number of principal components corresponding to nonzero eigenvalues obtained using kernel PCA on d dimensional dataset cannot be more than d.

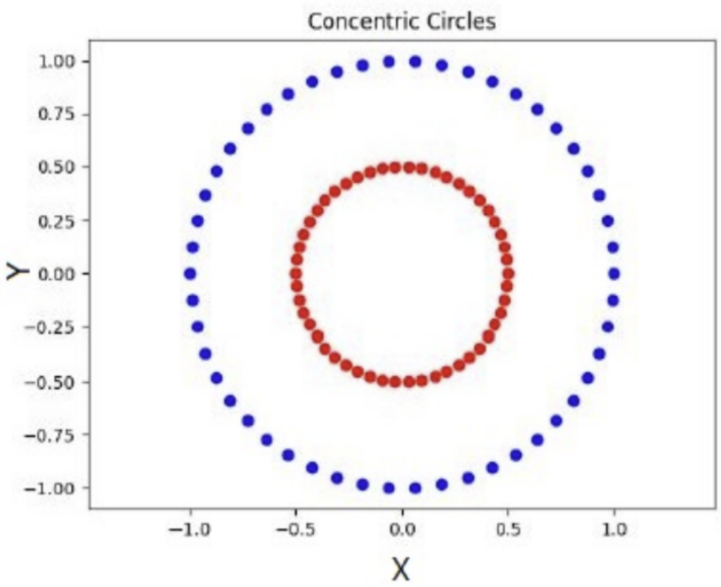
6406531929908. ✔ For every transformation mapping $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$, there exists a valid kernel function. Fact

6406531929909. ✔ For two valid kernels $k_1, k_2 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, $ak_1 + bk_2$ is a valid kernel function. Here both a and b are positive real numbers. part of some question in w2

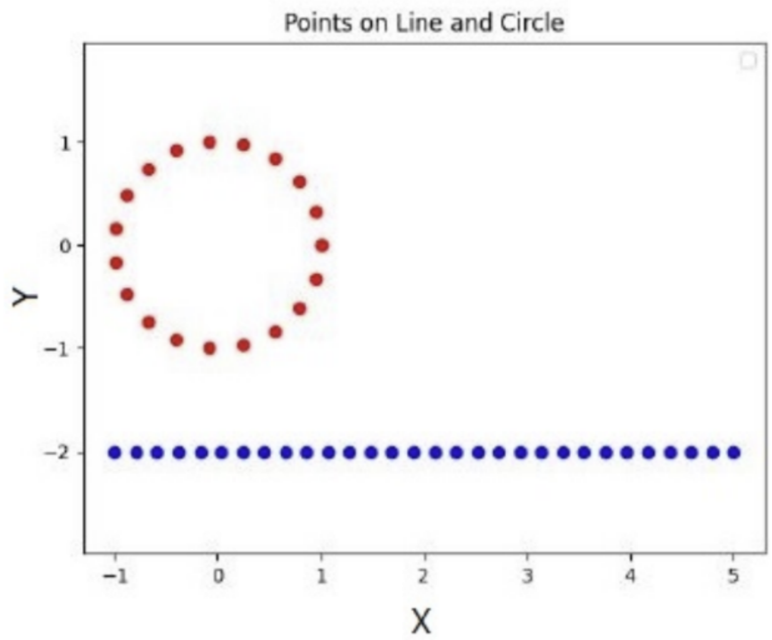
6406531929910. ✖ The output of a valid kernel function can never be a negative real number.

Which of the following clusters cannot be obtained as final clusters using Lloyd's algorithm? Assume that Lloyd's algorithm was run on these datasets with $k = 2$.

Options :

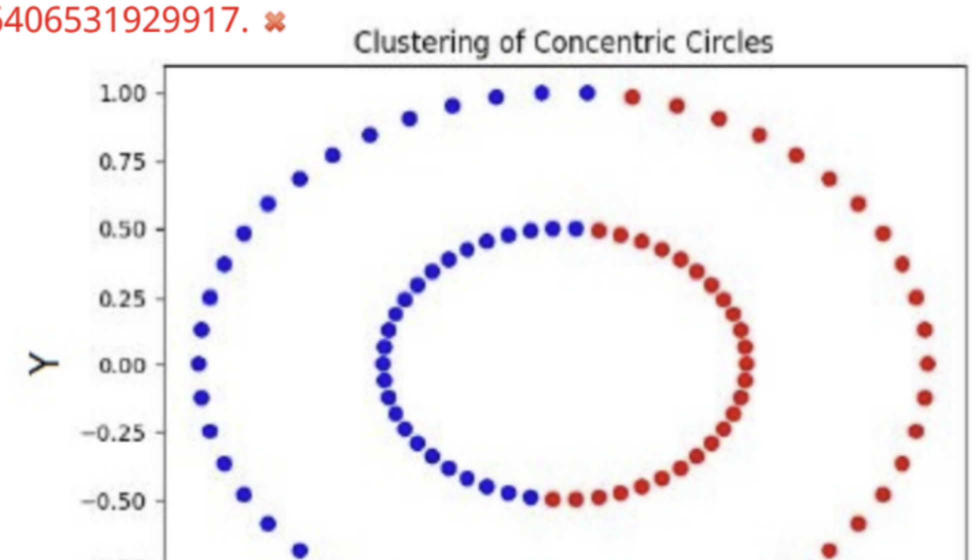


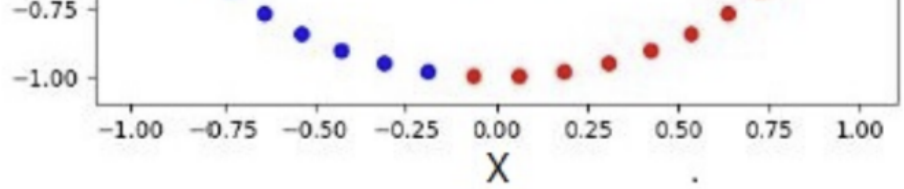
6406531929915. ✔

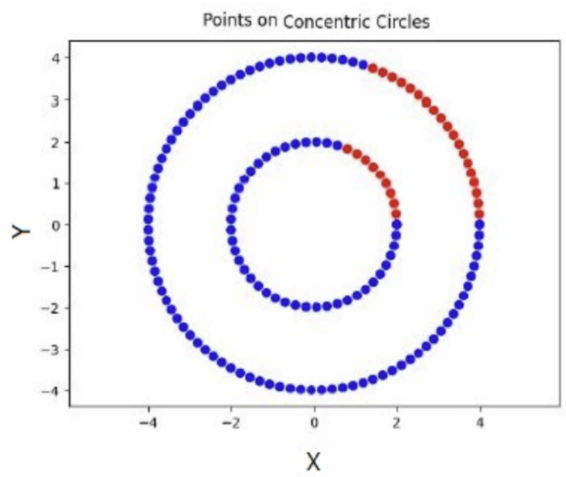


6406531929916. ✔

6406531929917. ✖







6406531929918. ✔

Which of the following statements is/are true?

Options :

6406531929911. ✔ $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, k(x_1, x_2) = (x_1^T x_2)^3$ is a valid kernel.

6406531929912. ✖ $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, k(x_1, x_2) = -(x_1^T x_2)^3$ is a valid kernel.

6406531929913. ✔ nonzero Eigenvalues of XX^T and X^TX are the same. fact

6406531929914. ✖ Eigenvectors of XX^T are same as eigenvectors of X^TX.

$k(x_1, x_2) = -(x_1^T x_2)^3$ (if $x_1 = [a, b]$, $x_2 = [c, d]$)

$\Rightarrow -\left[[a, b] \begin{bmatrix} c \\ d \end{bmatrix} \right]^3$

$\Rightarrow -\left[ac + bd \right]^3$

$\Rightarrow -\left[a^3c^3 + b^3d^3 + 3a^2c^2bd + 3b^2d^2ac \right]$

$\Rightarrow -a^3c^3 - b^3d^3 - 3a^2c^2bd - 3b^2d^2ac$

$\Rightarrow \phi(x_1^T) = [-a^3, -b^3, -3a^2b, -3b^2a]$

if we multiply $\phi(x_1^T)^T$ with $\phi(x_2^T)$ the values will be positive, which means $k(x_1, x_2) \neq \phi(x_1^T)^T \phi(x_2^T)$ cuz k is negative

Consider a dataset with 60 points in which all points are either 0 or 1. We use a Bernoulli distribution with the parameter p to model this problem. The prior and posterior distributions for the parameter p are Beta(10,5) and Beta(30,45) respectively. How many data points have the value 0 in this dataset? Note that $p = P(x = 1)$ as usual.

Prior = Beta(10, 5)

Posterior = Beta(30, 45)

$\Rightarrow$ Prior $\sim$ Beta(10, 5)

This is the probability <u>before</u> accounting for the actual distribution i.e. 0s and 1s.

Posterior $\sim$ Beta(30, 45)

This is the probability <u>after</u> accounting for the actual dataset, i.e. 0s and 1s.

To go from prior beta to posterior beta, we add number of 1s to the α and 0s to β. (for Beta(α, β))

$\Rightarrow$ No. of zeroes $= 45 - 5$

$\Rightarrow \underline{\underline{40}}$

Consider a GMM with two components:

$$\pi_1 = 0.3, \quad \pi_2 = 0.7$$

$$\mu_1 = 0, \quad \sigma_1^2 = 1$$

$$\mu_2 = 1, \quad \sigma_2^2 = 2$$

What is the probability that the point $x = 1.2$ comes from the first component? Use the following table for your reference. Here, $\mathcal{N}(x, \mu, \sigma^2)$ is the density of the Gaussian with mean μ and variance σ^2 evaluated at x.

Ans $\rightarrow$ Range(0.15, 0.35)

μ	σ^2	$\mathcal{N}(1.2, \mu, \sigma^2)$
0	1	0.194
0	2	0.197
1	1	0.391
1	2	0.279

Find: $P(z_i = 1 \mid x = 1.2)$

$$\Rightarrow P(z_i = 1 \mid x = 1.2) = \frac{P(x = 1.2 \mid z_i = 1) \cdot P(z_i = 1)}{P(x = 1.2)}$$

$$\Rightarrow \frac{0.194 \times 0.3}{0.3 \times 0.194 + 0.7 \times 0.279}$$

$\Rightarrow$ 0.0582

$$\Rightarrow \frac{0.0582}{0.0582 + 0.1953}$$

$$\Rightarrow \frac{0.0582}{0.2535} = \underline{\underline{0.229}}$$

Standard PCA has been performed on a centered dataset in $\mathbb{R}^3$. The first two principal components are given below:

$$\mathbf{w}_1 = \frac{1}{\sqrt{5}} \cdot \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{6}} \cdot \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

Consider the following data-point in the dataset:

$[1 \quad 2 \quad 1]^T$. (a, b) is the representation of this point in the coordinate system formed by the two principal components given above. The first and second coordinates correspond to PC-1 and PC-2 respectively.

Q) What is a?

Ans $\rightarrow$ Range(1.25, 1.5)

Q) What is b?

Ans $\rightarrow$ Range (0.25, 0.55)

Consider the following dataset in $\mathbb{R}$:

$$\{-3, \quad -1, \quad 0, \quad 1, \quad 4, \quad 299, \quad 300, \quad 301\}$$

Lloyd's algorithm ($K = 2$) is run on this dataset with the points -10 and 310 as the initial cluster centers. Let a, b be the final cluster centers on convergence, what is the value of the product ab?

$-3, -1, 0, 1, 4$ is assigned to cluster with mean $= -10$

$299, 300, 301$ is assigned to cluster with mean $= 310$

$$\mu_1 = a = \frac{4+1+0+-1+-3}{5} = \frac{1}{5}$$

$$\mu_2 = b = \frac{299+300+301}{3} = 300$$

$$a \times b = 1/5 \times 300 = \underline{\underline{60}}$$

Consider a dataset that has 100 data-points, each of which is either a zero or one. The Bernoulli distribution is used to model this data. If the MLE estimate for the parameter p of the Bernoulli distribution is 0.3, how many zeros does the dataset have?

$$\hat{p} = 0.3$$

$$\Rightarrow \frac{\sum x_i}{n} = 0.3$$

$$\Rightarrow \sum_{i=1}^{n} x_i = 0.3 \times 100 \quad [n = 100]$$

$$\Rightarrow \sum_{i=1}^{100} x_i = 30$$

$\Rightarrow$ There are 30 ones in the dataset.

$\Rightarrow$ No. of zeroes $= 100 - 30 = \underline{70}$

Set D2, Feb 26, 2023

Standard-PCA is performed on a centered dataset in $\mathbb{R}^3$. Two principal components are given below:

$$\frac{1}{2}\cdot\begin{bmatrix}1\\\sqrt{2}\\-1\end{bmatrix},\quad \frac{1}{\sqrt{2}}\cdot\begin{bmatrix}1\\0\\1\end{bmatrix}$$

Which of the following could be the third?

The principal components form an orthogonal basis.

$\Rightarrow \quad u_i^T u_j = 0 \qquad (i \neq j)$

a) $\frac{1}{2}\cdot\begin{bmatrix}1\\-\sqrt{2}\\-1\end{bmatrix}$

6406531562219. ✔

b) $\begin{bmatrix}0\\0\\0\end{bmatrix}$

6406531562220. ✖

c) $\frac{1}{\sqrt{2}}\cdot\begin{bmatrix}-1\\0\\1\end{bmatrix}$

6406531562221. ✖

d) $\frac{1}{\sqrt{3}}\cdot\begin{bmatrix}\sqrt{2}\\-1\\0\end{bmatrix}$

6406531562222. ✖

$$\begin{bmatrix}1/2\\1/\sqrt{2}\\-1/2\end{bmatrix},\quad \begin{bmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{bmatrix}$$

a) ✓

$$\begin{bmatrix}1/2 & -1/\sqrt{2} & -1/2\end{bmatrix}\begin{bmatrix}1/2\\1/\sqrt{2}\\-1/2\end{bmatrix}$$

$$\Rightarrow \frac{1}{4} - \frac{1}{2} + \frac{1}{4} = 0$$

$$\begin{bmatrix}1/2 & -1/\sqrt{2} & -1/2\end{bmatrix}\begin{bmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{bmatrix}$$

$$\Rightarrow \frac{1}{2\sqrt{2}} + 0 - \frac{1}{2\sqrt{2}} = 0$$

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

Options :

6406531562223. ✔ Yes

6406531562224. ✖ No

Let X be the data matrix of shape $d \times n$ with $d > n$ for a centered dataset. The eigenvector corresponding to the largest eigenvalue λ of X^TX is α_1. What will be the first principal component of the dataset?

Largest eigenvalue $= \lambda$

eigenvector $= \alpha_1$

$$\alpha_k = \frac{\beta_k}{\sqrt{n\lambda}}$$

$$\Rightarrow \quad \alpha = \frac{\alpha_1}{\sqrt{\lambda}} \qquad [\beta = \alpha_1]$$

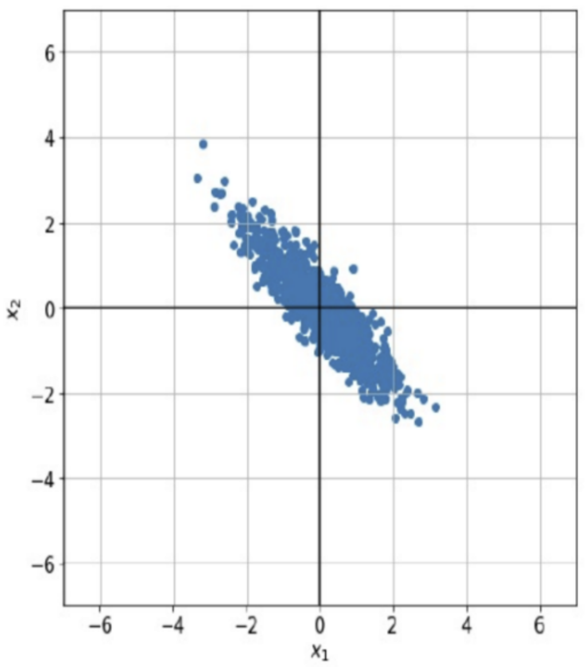
Projected first principal component $= X(\alpha) = \underline{\underline{\dfrac{X\alpha_1}{\sqrt{\lambda}}}}$

For any desired transformation $\phi(x)$, we can design a kernel function $k(x_1, x_2)$ that will evaluate $\phi(x_1)^T\phi(x_2)$.

Options :

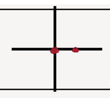
6406531562229. ✔ TRUE

6406531562230. ✖ FALSE

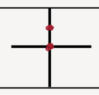


Which of the following could be the first principal component? Recall that the first P.C is the most important one.

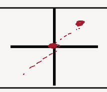
6406531562231. ✖ $\begin{bmatrix}1\\0\end{bmatrix}$



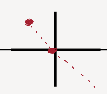
6406531562232. ✖ $\begin{bmatrix}0\\1\end{bmatrix}$



6406531562233. ✖ $\frac{1}{\sqrt{2}} \cdot \begin{bmatrix}1\\1\end{bmatrix}$



6406531562234. ✔ $\frac{1}{\sqrt{2}} \cdot \begin{bmatrix}-1\\1\end{bmatrix}$



Only 4th option lies along the dataset.

In standard PCA, which of the following are correct formulations of the optimization problem? The dataset $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, is centered, each point lies in $\mathbb{R}^d$, and $\mathbf{C}$ is the covariance matrix. $\mathbf{w} \in \mathbb{R}^d$ is the variable that we are optimizing over.

Options :

6406531562235. ✔

$$\min_{\mathbf{w}} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{w})\mathbf{w}||^2$$
$$\text{subject to} \quad ||\mathbf{w}|| = 1$$

→ minimizing error

$$\max_{\mathbf{w}} \quad \mathbf{w}^T\mathbf{C}\mathbf{w}$$

→ maximizing variance

6406531562236. ✔ subject to $||\mathbf{w}|| = 1$

$$\max_{\mathbf{w}} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

6406531562237. ✖ subject to $||\mathbf{w}|| = 1$

$$\min_{\mathbf{w}} \quad \mathbf{w}^T \mathbf{C} \mathbf{w}$$

6406531562238. ✖ subject to $||\mathbf{w}|| = 1$

Consider a GMM with 3 components that is used to model a dataset that has 100 points. The EM algorithm is run on this dataset to estimate the parameters of the GMM. After convergence, some of the values of λ_k^i for the specific point $\mathbf{x}_{10}$ are given below:

$$\lambda_1^{10} = 0.2, \quad \lambda_2^{10} = 0.5$$

Select all true statements.

Options :

6406531562239. ✖ If a point is picked randomly from the dataset of 100 points, then there is 20% chance that it comes from the first component.

$(0.2) \quad (0.5) \quad (?)$

$\lambda_1^{10} + \lambda_2^{10} + \lambda_3^{10} = 1$

6406531562240. ✔ $\lambda_3^{10} = 0.3$

6406531562241. ✖ $\sum_{i=1}^{100} \lambda_3^i = 1$ λ_3 for all 'i's will be the same

6406531562242. ✔ There is a 50% chance that the point $\mathbf{x}_{10}$ comes from the second component.

Consider a dataset with n points and a GMM with K components. If we fix λ and maximize for θ, what is our estimate for the mean of the k^{th} component?

Options :

6406531562243. ✔ It is the weighted mean of the n points, where the weight for point i in component k is given by λ_k^i

6406531562244. ✖ It is the mean of the n points.

6406531562245. ✖ $\hat{\mu}_k = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

6406531562246. ✔ $\hat{\mu}_k = \frac{\sum_{i=1}^{n} \lambda_k^i x_i}{\sum_{i=1}^{n} \lambda_k^i}$

With respect to the Lloyd's algorithm, choose the correct statements:

With respect to the Lloyd's algorithm, choose the correct statements.

Options :

6406531562247. ✔ The partition configurations cannot repeat themselves.

6406531562248. ✖ After doing the reassignments (consider at least one point reassigned to the new cluster), we might get the same means for all clusters.

6406531562249. ✔ Objective function after making the re-assignments strictly reduces.

6406531562250. ✖ Objective function after making the re-assignments may increase.

6406531562251. ✔ A change in the objective function's value indicates that the partition configuration has changed.

6406531562252. ✖ For partitioning n data points across k partitions, Lloyd's algorithm takes k^n iterations to converge.

Consider a dataset of 20 points where the i^{th} data-point is given by:

$$\mathbf{x}_i = a_i \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix} + b_i \cdot \begin{bmatrix} -1 \\ 1 \\ 3 \\ 0 \end{bmatrix} \quad ?$$

where, a_i and b_i are real numbers such that $\sum_{i=1}^{20} a_i = \sum_{i=1}^{20} b_i = 0$. Standard PCA is performed on this dataset. If the top two principal components are retained and used to reconstruct the dataset, what is the reconstruction error?

Hint: Think about what happens in $\mathbb{R}^2$ or $\mathbb{R}^3$ for a similar situation and extend this idea to $\mathbb{R}^4$.

Consider the following prior for the parameter p of a Bernoulli distribution:

$$p \sim \text{Beta}(3, 2)$$

The dataset observed is as follows:

$$\{1,\quad 0,\quad 1,\quad 0,\quad 1,\quad 1,\quad 1,\quad 0,\quad 0,\quad 0,\quad 1,\quad 1,\quad 1\}$$

What is $\hat{p}$, a point estimate for the parameter of the Bernoulli distribution, if we use the expectation of the posterior as the method of estimation? Enter your answer correct to three decimal places.

posterior $\sim$ Beta$(11, 7)$

$$E[\text{posterior}] = \frac{\alpha}{\alpha+\beta} = \frac{11}{18} = \underline{\underline{0.611}}$$

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?

$d = 3$

$p = 2$

new dimension $= {}^{d+p}C = {}^{5}C = 10$

new dimension =

Consider a centered dataset in $\mathbb{R}^{100}$ that has 1000 data-points. Call this dataset D_1. Standard PCA is performed on D_1 and the **scalar** projections of D_1 on the top 5 principal components are computed. Call the resulting dataset D_2.

NOTE: D_2 is only made up of the scalar projections. All the principal components are thrown out after computing the scalar projections and are not a part of D_2.

Each data-point in D_2 belongs to $\mathbb{R}^k$. What is the value of k?

$k = 5$

If each real number occupies unit storage space in the memory, compute:

$$\frac{\text{size}(D_1)}{\text{size}(D_2)}$$

$$\frac{\text{size}(D_1)}{\text{size}(D_2)} = \frac{1000 \times 100}{1000 \times 5} = 20$$

Assume that you have a dataset of five points $\{x_1, x_2, x_3, x_4, x_5\}$, all of which are non-negative. You hypothesise that the data points are iid random variables with the following density:

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

What is the log-likelihood of this dataset under this distribution? $\ln$ represents the natural logarithm or $\log_e$.

Options :

6406531562258. ✖ $\prod_{i=1}^{5} \lambda e^{-\lambda x_i}$

6406531562259. ✖ $\sum_{i=1}^{5} \lambda e^{-\lambda x_i}$

6406531562260. ✔ $\sum_{i=1}^{5} [\ln(\lambda) - \lambda x_i]$

6406531562261. ✖ $\prod_{i=1}^{5} [\ln(\lambda) - \lambda x_i]$

You are given the actual values of these observations:

$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3, \quad x_4 = 4, \quad x_5 = 5$

What is the maximum likelihood estimate for λ? Enter your answer correct to three decimal places.

$$\frac{d}{d\lambda}\sum_{i=1}^{5}\left[\ln(\lambda) - \lambda x_i\right] = 0$$

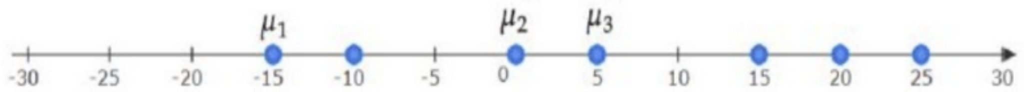
$$\Rightarrow \sum_{i=1}^{5}\left[1/\lambda - x_i\right] = 0$$

$$\Rightarrow \frac{n}{\lambda} = \sum x_i$$

$$\Rightarrow \frac{n}{\sum x_i} = \hat{\lambda}$$

$$\therefore \hat{\lambda} = \frac{5}{5+4+3+2+1} = \frac{5}{15} = \frac{1}{3} = \underline{\underline{0.333}}$$

Consider the following one-dimensional dataset of seven points that are distributed as follows:



k-means algorithm with $k = 3$ was run on the given data points. μ_1, μ_2, and μ_3 are the initial cluster means.

How many points belong to cluster 3 (mean μ_3) for the initial clusters?

4 cuz (5, 15, 20, 25) is closest to initial cluster mean 5.

What will be the mean μ_3 after 1^{st} iteration? Enter your answer correct to two decimal places.

$$\mu = \frac{5+15+20+25}{4} = \frac{65}{4} = 16.25$$

How many points belong to cluster 2 (mean μ_2) after the 1^{st} iteration?

After first iteration, new means are

$\mu_1 = -\frac{15-10}{2} = -12.5$, $\mu_2 = \frac{0}{2} = 0$, $\mu_3 = \frac{25+20+15+5}{2} = 32.5$

We can see that now with new means,

5 is closer to μ_2, so it will go to μ_2.

$\therefore$ Cluster 2 has $\underline{\underline{2}}$ points after first iteration.

How many iterations will it take for the algorithm to converge with the given initial clusters?

2

What will be the final cluster means?

Options :

6406531562268. ✖ $\mu_1 = -15, \ \ \mu_2 = 0, \ \ \mu_3 = 5$

6406531562269. ✖ $\mu_1 = -10, \ \ \mu_2 = 0, \ \ \mu_3 = 15$

6406531562270. ✖ $\mu_1 = -12.5, \ \ \mu_2 = 0, \ \ \mu_3 = 16.25$

6406531562271. ✖ $\mu_1 = -12.5, \ \ \mu_2 = 2.5, \ \ \mu_3 = 16.25$

6406531562272. ✔ $\mu_1 = -12.5, \ \ \mu_2 = 2.5, \ \ \mu_3 = 20$

A k-means++ algorithm with k = 3 was applied to following 2D points:

(0,0),(1,2),(3,1),(4,7),(-1,9),(4,-2)

(0,0) is chosen as the first cluster mean.

Which point has the highest probability of being chosen as the 2^{nd} cluster mean? Use the manhattan distance to compute the distances.

Manhattan distance $= |x_1 - y_1| + |x_2 - y_2|$

Options :

6406531562273. ✖ (3,1)

6406531562274. ✔ (4,7)

6406531562275. ✖ (-1,9)

6406531562276. ✖ (4,-2)

If the point with the highest score is chosen as the 2^{nd} cluster mean(answer from previous

question), Which point has the highest probability of being chosen as 3^{rd} the cluster mean? Use the manhattan distance to compute the distances.

Options :

6406531562277. ✖ (3,1)

6406531562278. ✖ (4,7)

6406531562279. ✔ (-1,9)

6406531562280. ✖ (4,-2)

FN, Oct 16, 2022

Consider the following kernel:

$$k : R^2 \times R^2 \to R$$
$$k(x, y) = (x^T y)^2$$

Which of the following transformation mapping ϕ may correspond to the kernel k?

have done this many times already 😑

Options :

6406531285582. ✖ $\phi([x_1, x_2]^T) = [x_1, x_1x_2, x_2]^T$

6406531285583. ✖ $\phi([x_1, x_2]^T) = [x_1^2, x_1 + x_2, x_2^2]^T$

6406531285584. ✖ $\phi([x_1, x_2]^T) = [x_1, \sqrt{2}x_1^2x_2^2, x_2]^T$

6406531285585. ✔ $\phi([x_1, x_2]^T) = [x_1^2, \sqrt{2}x_1x_2, x_2^2]^T$

A function k is defined as

$$k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$
$$k(x_1, x_2) = x_1x_2 + x_1^2x_2^2 + \underbrace{x_1^3x_2}_{} + 1$$

Is k a valid kernel?

this term doesn't have matching x_1 and x_2, i.e. powers of x_1 and x_2 should be same.

Options :

6406531285586. ✖ Yes

6406531285587. ✔ No

Let X be a data matrix of shape (d, n) for a centered dataset. The first principal component of the dataset is $\left[\frac{1}{2}, \frac{\sqrt{3}}{2}\right]^T$. What will the scalar proxy of the point $[1, 2]^T$ be on the first principal component?

$$[1, 2]\left[\frac{1}{2}, \frac{\sqrt{3}}{2}\right] = \text{proxy/constant}$$

$$\Rightarrow \quad \frac{1}{2} + \frac{2\sqrt{3}}{2}$$

Options :

6406531285588. ✖ $1 + 2\sqrt{3}$

6406531285589. ✔ $\frac{1 + 2\sqrt{3}}{2}$

6406531285590. ✖ $\frac{1 + 2\sqrt{3}}{4}$

6406531285591. ✖ $\frac{1 + 2\sqrt{3}}{\sqrt{5}}$

Consider the scenario where you observe that 8 of your friends $\{f_1, \ldots, f_8\}$ have scored the following marks in a quiz: $\{6, 3, -5, -4, 2, -3, 5, -2\}$ respectively.

Do this question using a number line and calc the new means.

Which of the cluster initializations given below will result in clusters where your friends with positive marks are in one cluster and the rest are in another after executing one step of the Lloyd's algorithm?

$I_1 : z_1 = z_2 = z_3 = z_4 = z_5 = z_6 = z_7 = 2$ and $z_8 = 1$
$I_2 : z_1 = 1, z_2 = z_3 = z_4 = z_5 = z_6 = z_7 = z_8 = 2$

Options :

6406531285595. ✖ I_1: No, I_2: No

6406531285596. ✖ I_1: Yes, I_2: Yes

6406531285597. ✔ I_1: Yes, I_2: No

6406531285598. ✖ I_1: No, I_2: Yes

6406531285599. ✖ Insufficient Information

Question Label : Multiple Choice Question

Consider the following data set:

$$\left\{x_1=\begin{bmatrix}0\\2\end{bmatrix}, x_2=\begin{bmatrix}2\\0\end{bmatrix}, x_3=\begin{bmatrix}0\\0\end{bmatrix}, x_4=\begin{bmatrix}0\\-2\end{bmatrix}, x_5=\begin{bmatrix}-2\\0\end{bmatrix}, x_6=\begin{bmatrix}0\\1\end{bmatrix}, x_7=\begin{bmatrix}0\\-1\end{bmatrix}\right\}$$

For $k=3$, assume that the following cluster assignment is given to us which shows the clusters assigned to all data points but x_3 :

$$z=\{1,2,?,3,2,1,3\}$$

What should be the cluster assigned to the data point x_3?

Options :

6406531285592. ✖ 1

6406531285593. ✔ 2

6406531285594. ✖ 3

$$\mu_1 = \frac{\begin{bmatrix}0\\2\end{bmatrix}+\begin{bmatrix}0\\1\end{bmatrix}}{2} = \begin{bmatrix}0\\1.5\end{bmatrix}$$

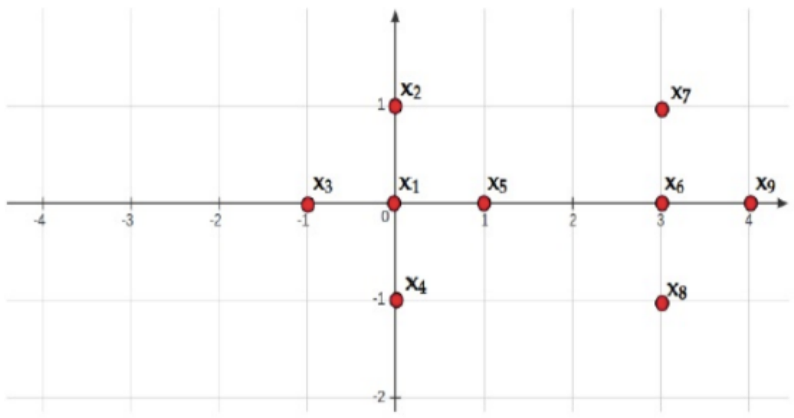
$$\mu_2 = \frac{\begin{bmatrix}2\\0\end{bmatrix}+\begin{bmatrix}-2\\0\end{bmatrix}}{2} = \begin{bmatrix}0\\0\end{bmatrix}$$

$$\mu_3 = \frac{\begin{bmatrix}0\\-2\end{bmatrix}+\begin{bmatrix}0\\-1\end{bmatrix}}{2} = \begin{bmatrix}0\\-1.5\end{bmatrix}$$

x_3 is closest to μ_2

$\Rightarrow$ It will be assigned to $\underline{\underline{2^{nd}}}$ cluster.

Consider the data points shown in the following image:



?

How to find objective value?

$$\mu_1 = \frac{\begin{bmatrix}0\\0\end{bmatrix}+\begin{bmatrix}1\\0\end{bmatrix}+\begin{bmatrix}-1\\0\end{bmatrix}+\begin{bmatrix}0\\1\end{bmatrix}+\begin{bmatrix}0\\-1\end{bmatrix}}{5} = \begin{bmatrix}0\\0\end{bmatrix}$$

$$\mu_2 = \frac{\begin{bmatrix}3\\0\end{bmatrix}+\begin{bmatrix}4\\0\end{bmatrix}+\begin{bmatrix}3\\1\end{bmatrix}+\begin{bmatrix}3\\-1\end{bmatrix}}{4} = \begin{bmatrix}3.25\\0\end{bmatrix}$$

Consider that in the initialization step of K-means with $k=2$, the data points x_1 and x_6 got selected as initial cluster centers. That is, $\mu_1^0=(0,0)$ and $\mu_2^0=(3,0)$. As per these cluster centers, the data points were then assigned to either cluster 1 or cluster 2. After this assignment, what will be the value of the objective function (Obj)? Further, as per the assignment of data points, the cluster centers will be re-computed for the next iteration. What will be the value of μ_1^1 and μ_2^1?

Options :

6406531285600. ✖ $Obj=17, \mu_1^1=(1,0), \mu_2^1=(3.33,0)$

6406531285601. ✖ $Obj=17, \mu_1^1=(1,0), \mu_2^1=(3,0)$

6406531285602. ✖ $Obj=7, \mu_1^1=(0,0), \mu_2^1=(3.33,0)$

6406531285603. ✔ $Obj=7, \mu_1^1=(0,0), \mu_2^1=(3.25,0)$

In the context of EM algorithm, select all true statements from the options given below. There are n data-points and K mixtures. The index i corresponds to the i^{th} data-point, the index k corresponds to the k^{th} mixture. f is the density of a Gaussian. All other symbols have their usual meaning.

Already done many times

Options :

6406531285604. ✔ $\pi_k=P(z_i=k)$

6406531285605. ✖ $\sum_{i=1}^{n} \lambda_k^i = 1$

6406531285606. ✔ $\lambda_k^i = P(z_i = k \mid x_i)$

6406531285607. ✖ $\lambda_k^i = f(x_i \mid z_i = k)$

Consider a centered dataset of 100 points in $\mathbb{R}^5$. Standard PCA is performed on this dataset. Consider the following ratio:

$$\theta = \frac{\text{variance along the first principal component}}{\text{sum of variances along all five principal components}}$$

What is the minimum value of θ?

Ans $\longrightarrow$

Range(0.19, 0.21)

Only when all variances are equal, the first principal component variance will be minimum.

$\Rightarrow \theta = \frac{x}{x+x+x+x+x} = \frac{x}{5x} = 0.2$ [x = variance of all principal components]

Consider a centered dataset of 1000 points in $\mathbb{R}^{10}$. Standard PCA is performed on this dataset. What is the maximum number of principal components that can be retained so that the compression ratio does not fall below 1.4 ? The compression ratio is defined as:

$$\frac{\text{size of original dataset}}{\text{size of reconstructed dataset}}$$

Assume that each real number occupies one unit of storage space. Note that we are interested in the reconstructions and not just in the scalar projections. So, the reconstructed dataset will also be in $\mathbb{R}^{10}$.

Compression ratio = $\frac{\text{size of original}}{\text{size of reconstructed}}$

$\Rightarrow 1.4 = \frac{1000 \times 10}{1000 \times x}$

$\Rightarrow x = \frac{10}{1.4} \simeq 7$

Consider a dataset with 60 points in which all points are either 0 or 1. We use a Bernoulli distribution with parameter p to model this problem. The prior and posterior distributions for the parameter p are Beta(10, 5) and Beta(30, 45) respectively. How many data-points have the value 0 in this dataset? Note that $p = P(x = 1)$ as usual.

Similar question done above.

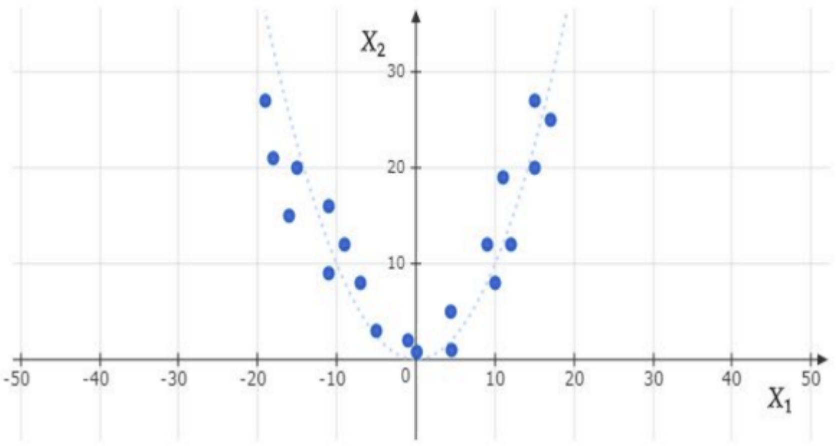
A GMM is fit for a dataset with 5 points. At some time-step in the EM algorithm, the following are the values of λ_k^i for all points in the dataset for the k^{th} mixture after the E-step:

$$\lambda_k^1 = 0.9$$
$$\lambda_k^2 = 0.8$$
$$\lambda_k^3 = 0.7$$
$$\lambda_k^4 = 0.6$$
$$\lambda_k^5 = 0.5$$

What is the estimate of π_k after the M-step?

$\hat{\pi}_k^{MML} = \sum_{i=1}^{n} \lambda_k^i = \frac{0.9+0.8+0.7+0.6+0.5}{} = 0.7$

Consider the dataset as shown in the figure. We want to project the dataset into another feature space so that it lives in a linear subspace of a higher-dimension space. We apply the kernel PCA with a polynomial kernel of the appropriate degree to achieve the same. What will be the dimension of transformed feature space?



From the graph we can see that this graph is similar to x^2 graph

$\Rightarrow$ Kernel PCA is applied.

To convert quadratic to linear we use the following map,

if $x = [f_1 \quad f_2]$

$\Rightarrow \phi(x) = [1 \quad f_1^2 \quad f_2^2 \quad f_1 f_2 \quad f_1 \quad f_2]$

$\therefore$ Dimension of transformed feature space $= \underline{\underline{6}}$

Another way to look at this, on graph it can be seen that dataset belongs to R^2 as there are two variables (X_1, X_2)

$\Rightarrow d = 2$

The datapoints lie on the second degree curve, (x^2)

$\Rightarrow p = 2$

$\therefore {}^{d+p}C_d = {}^{2+2}C_2 = {}^4C_2 = \underline{\underline{6}}$

Consider the following data set:

$$\left\{ x_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, x_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, x_4 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, x_5 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \right\}$$

With $k = 2$, as per k-means++,

(Enter your answers correct up to three decimal places)

What is the probability of the following points x_2, x_1 (in that order) being chosen as initial cluster centers?

Ans $\rightarrow$ Range (0.03, 0.05)

In K-Means++ first mean is chosen randomly from the dataset,

$\Rightarrow$ Probability that any point is chosen $= 1/n = 1/5$ $[n = 5]$

After initialization, points are chosen according to the score.

Distance of points from mean (x_2)

$x_1 x_2 = (0-2)^2 + (2-0)^2 = 8$

$x_3 x_2 = (0-2)^2 + (0-0)^2 = 4$

$x_4 x_2 = (0-2)^2 + (-2-0)^2 = 8$

$x_5 x_2 = (2--2)^2 + (0-0)^2 = 16$

Probability of x_1 being chosen $= \dfrac{8}{8+4+8+16} = \dfrac{\cancel{8}^{\times 2}}{\cancel{36}_{\times 9}} = \dfrac{2}{9}$

Probability of x_2 and x_1 being chosen $= \dfrac{1}{5} \times \dfrac{2}{9} \simeq \underline{0.0444}$

What is the probability of the following points x_2, x_3 (in that order) being chosen as initial cluster centers? Ans → Range (0.01, 0.03)

What is the probability of the following points x_2, x_5 (in that order) being chosen as initial cluster centers? Ans → Range (0.08, 0.095)

Consider a dataset that has 4 points in $\mathbb{R}^2$ that lie on a line passing through the origin:

$$D = \left\{ \begin{bmatrix} -2 \\ -6 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}$$

Based on the above data, answer the given subquestions.

Is this dataset centered?

Options :

6406531285616. ✔ Yes

6406531285617. ✖ No

Choose all representatives $\mathbf{w}$ for this dataset such that $||\mathbf{w}|| = 1$.

Options :

6406531285618. ✔ $\frac{1}{\sqrt{10}} \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

6406531285619. ✔

$\frac{1}{\sqrt{10}} \cdot \begin{bmatrix} -1 \\ -3 \end{bmatrix}$

6406531285620. ✖ $\begin{bmatrix} -1 \\ -3 \end{bmatrix}$

6406531285621. ✖ $\begin{bmatrix} 2 \\ 6 \end{bmatrix}$

If standard PCA is performed on this dataset, what is the variance along the first principal component? Ans → 25

If standard PCA is performed on this dataset, what is the variance along the second principal component? ?

Ans $\rightarrow 0$

Assume that you have a dataset of four points $\{x_1, x_2, x_3, x_4\}$, all of which lie in $[0, 1]$. You hypothesise that the data points are iid random variables with the following density:

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Based on the above data, answer the given subquestions.

What is the log-likelihood of this dataset under this distribution

Options :

6406531285624. ✖ $\prod_{i=1}^{4} \theta x_i^{\theta-1}$

6406531285625. ✖ $\sum_{i=1}^{4} \theta x_i^{\theta-1}$

6406531285626. ✔ $\sum_{i=1}^{4} [\log\theta + (\theta-1)\log x_i]$

6406531285627. ✖ $\prod_{i=1}^{4} [\log\theta + (\theta-1)\log x_i]$

You are given the actual values of these observations:

$$x_1 = \frac{1}{e}, \quad x_2 = \frac{1}{e^2}, \quad x_3 = \frac{1}{e^3}, \quad x_4 = \frac{1}{e^4}$$

What is the maximum likelihood estimate for θ? Use $\log_e$ wherever appropriate.

Hint: Differentiate the log-likelihood with respect to θ and set it to zero.

$$\frac{d}{d\theta}\left(\sum_{i=1}^{n} [\log\theta + (\theta-1)\log x_i\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left[\frac{1}{\theta} + \log x_i\right] = 0$$

$$\Rightarrow \frac{n}{\theta} = -\sum_{i=1}^{n} \log x_i$$

$$\Rightarrow \frac{-n}{\sum_{i=1}^{n} \log x_i} = \hat{\theta}$$

$$\Rightarrow \hat{\theta} = \frac{-4}{-1 + -2 + -3 + -4} = \frac{4}{10} = 0.4$$