Giorgos Stylianos

#3352410

Homework 6

1) Correct   3,5
2) Correct   1, 3
3) Correct   3, 4

4) Correct   4
5) Correct   3
6) Correct   4
7) Correct   1, 2.
8) Correct   2
9) Correct   1, 4, 5, 7
10) Correct   2, 3, 4
11) Correct   4
12) Correct   2

Exercise 13

• Step1 Likelihood function

$$p(X|\vartheta) = \prod_{i=1}^{N} p(x_i|\vartheta) - \prod_{i=1}^{N} (2\vartheta x_i \exp(-\vartheta x_i^2))$$

$$p(X|\vartheta) = 2^N \vartheta^N \left(\prod_{i=1}^{N} x_i\right) \exp\left(-\vartheta \sum_{i=1}^{N} x_i^2\right)$$

• Step 2 Prior Distribution

$$p(\vartheta) = \frac{1}{\sqrt{2\pi 60^2}} \exp\left(-\frac{(\vartheta-\vartheta_0)^2}{2 60^2}\right)$$

• Step 3 Posterior

$$\log p(\vartheta|x) = N\ln\vartheta - \vartheta \sum_{i=1}^{N} x_i^2 - \frac{(\vartheta-\vartheta_0)^2}{2 60^2}$$

• Step 4 Maximize the posterior

$$\frac{d}{d\vartheta} \log p(\vartheta|x) = \frac{N}{\vartheta} - \sum_{i=1}^{N} x_i^2 - \frac{\vartheta-\vartheta_0}{60^2}$$

$$\frac{N}{\vartheta} - \sum_{i=1}^{N} x_i^2 - \frac{\vartheta-\vartheta_0}{60^2} = 0$$

$$\frac{N}{\vartheta} = \sum_{i=1}^{N} x_i^2 + \frac{\vartheta-\vartheta_0}{60^2}$$

$$N = \vartheta \sum_{i=1}^{N} x_i^2 + \frac{\vartheta^2}{60^2} - \frac{\vartheta\vartheta_0}{60^2}$$

$$\frac{1}{60^2}\vartheta^2 + \left(\sum_{i=1}^{N} x_i^2 - \frac{\vartheta_0}{60^2}\right)\vartheta - N = 0$$

$$\vartheta = \frac{1}{2}\left(\left(\vartheta_0 - \frac{60^2}{N}\sum_{i=1}^{N} x_i^2\right) + \sqrt{\left(\vartheta_0 - \frac{60^2}{N}\sum_{i=1}^{N} x_i^2\right)^2 + 4N 60^2}\right)$$

$$\vartheta_{MAP} = \frac{1}{2}\left(\left(\vartheta_0 - \frac{60^2}{N}\sum_{i=1}^{N} x_i^2\right) + \sqrt{\left(\vartheta_0 - \frac{60^2}{N}\sum_{i=1}^{N} x_i^2\right)^2 + 4N 60^2}\right)$$

b) i) $N \to \infty$

When $N \to \infty$ litelihood function is concentrated around the true value of θ. So MAP approaches ML which is derived from the data.

ii) As $\sigma_0^2 \to \infty$.

The prior variance σ_0^2 becomes large the prior becomes weak and it behaves almost like a non-informative prior. MAP approaches ML since the prior has a little effect.

iii) $\sigma_0^2 \to 0$.

As σ_0^2 becomes small prior $\to$ strong MAP is concentrated by the prior information. So θ_{MAP} will be close to θ_0.

d) ϑMAP for data.

$x_1 = 2, x_2 = 2.2, x_3 = 2.7, x_4 = 2.4, x_5 = 2.6$

$\vartheta_0 = 1.8$

$\sigma_0^2 = 1.$

$\sum_{i=1}^{5} x_i^2 = 2^2 + 2.2^2 + 2.7^2 + 2.4^2 + 2.6^2 = 28.65.$

$\sum_{i=1}^{5} x_i = 2 + 2.2 + 2.7 + 2.4 + 2.6 = 11.9$

$\hat{\vartheta}_{MAP} = \dfrac{28.65 + 1.8}{5 + 1} = \dfrac{30.45}{6} = 5.075.$

$p(2.3; 5.075) = 2.5.075.2.3 \exp(-26.84)$

$p(2.9; 5.075) = 2.5.075.2.9 \exp(-42.67).$

Exercise 14.

a) Likelihood is

$$P(Y|\lambda) = \prod_{i=1}^{N} \lambda e^{-\lambda x_i} = \lambda^N \prod_{i=1}^{N} e^{-\lambda x_i} = \lambda^N e^{-\lambda \sum_{i=1}^{N} x_i}$$

b) Product is

$$p(\lambda) p(Y|\lambda) = \frac{b^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda}$$

$$\frac{d A(\lambda)}{d\lambda} = 0 \implies \frac{b^\alpha}{\Gamma(\alpha)} \left[(\alpha+N-1)\lambda^{\alpha+N-2} e^{-\lambda(b+\sum_{i=1}^N x_i)} \right. $$

$$\left. - \lambda^{\alpha+N-1}\left(b + \sum_{i=1}^N x_i\right) e^{-\lambda(b+\sum_{i=1}^N x_i)} \right] =$$

$$\frac{b^\alpha}{\Gamma(\alpha)} \lambda^{\alpha+N-2} e^{-\lambda(b+\sum_{i=1}^N x_i)} \left[\alpha+N-1-\lambda(b+\sum_{i=1}^N x_i) \right] = 0 \iff$$

$$\alpha + N - 1 - \lambda(b + \sum_{i=1}^N x_i) = 0 \iff$$

$$\boxed{\lambda_{MAP} = \frac{\alpha + N - 1}{b + \sum_{i=1}^N x_i}}$$

c)

$$p(x) = \begin{cases} \dfrac{\alpha+N-1}{b + \sum_{i=1}^N x_i} e^{-\left(\frac{\alpha+N-1}{b+\sum_{i=1}^N x_i}\right)x}, & x \geq 0 \\[3mm] 0 & x < 0 \end{cases}$$

d) Posterior is

$$p(\lambda|Y) = \frac{p(\lambda)p(Y|\lambda)}{p(Y)} = \frac{1}{p(Y)} \frac{b^\alpha}{\Gamma(\alpha)} \lambda^{\alpha+N-1} e^{-\lambda(b+\sum_{i=1}^N x_i)}$$

e) Prior and posterior are both gamma distribution

f) $\int_0^\infty \lambda e^{-\lambda x} \dfrac{b'^{\alpha'}}{\Gamma(\alpha')} \lambda^{\alpha'-1} e^{-b'\lambda} \, d\lambda$

$= \dfrac{b'^{\alpha'}}{\Gamma(\alpha')} \int_0^\infty \lambda^{\alpha'} e^{-(x+b')\lambda} \, d\lambda$

$= \dfrac{b'^{\alpha'}}{\Gamma(\alpha')} \cdot \dfrac{\Gamma(\alpha'+1)}{(x+b')^{\alpha'+1}} = \dfrac{b'^{\alpha'}}{\Gamma(\alpha')} \cdot \dfrac{\alpha' \Gamma(\alpha')}{(x+b')^{\alpha'+1}} = \dfrac{\alpha' b'^{\alpha'}}{(x+b')^{\alpha'+1}}$

$\dfrac{\alpha' b'^{\alpha'}}{(x+b')^{\alpha'+1}}$ is a lomax distribution where $c=\alpha', d=b'$

g) It is $N=6$ and $\sum_{i=1}^N x_i = 15$

$p(x) = \begin{cases} \dfrac{\alpha+N-1}{b+\sum_{i=1}^N x_i} e^{-\left(\frac{\alpha+N-1}{b+\sum_{i=1}^N x_i}\right)x} & , x \geq 0 \\ \\ 0 & , x \leq 0. \end{cases}$

$= \begin{cases} \dfrac{7}{17} e^{-\left(\frac{7}{17}\right)x} & , x \geq 0 \\ \\ 0 & , x \leq 0. \end{cases}$

w) $\alpha' = \alpha + N = 2+6 = 8$, $b' = b\sum_{i=1}^N x_i = 2+15 = 17$.

$p(x|Y) = \begin{cases} \dfrac{8 \cdot 17^8}{(x+17)^{8+1}} & , x \geq 0 \\ \\ 0 & , x < 0 \end{cases} = \begin{cases} \dfrac{8 \cdot 17^8}{(x+17)^9} & , x \geq 0 \\ \\ 0 & , x < 0. \end{cases}$

i) $p(2.5) = \dfrac{7}{17} e^{-(7/17)2.5} = 0.1471$, $p(2.5|Y) = \dfrac{8 \cdot 17^8}{(2.5+17)^9} = 0.1369$

Exercise 15

Lagrangian is
$$L(\theta) = \sum_{i=1}^{N}(x_u - \theta)^2 + \lambda((\theta - \theta_0)^2 - \rho)$$
Constraint is $(\theta - \theta_0)^2 \leq \rho$.

Derivative with respect to θ.
$$\frac{d}{d\theta}\left[\sum_{i=1}^{N}(x_u - \theta)^2\right] = -2\sum_{u=1}^{N}(x_u - \theta)$$

$$\frac{d}{d\theta}\left[\lambda((\theta - \theta_0)^2 - \rho)\right] = 2\lambda(\theta - \theta_0)$$

$$\frac{dL(\theta)}{d\theta} = -2\sum_{i=1}^{N}(x_u - \theta) + 2\lambda(\theta - \theta_0) = 0.$$

$$\sum_{u=1}^{N}(x_u - \theta) = \lambda(\theta - \theta_0) \iff$$
$$\sum_{u=1}^{N}x_u - N\theta = \lambda(\theta - \theta_0).$$

$$N\theta + \lambda\theta = \sum_{u=1}^{N}x_u + \lambda\theta_0 \implies \theta(N+\lambda) = \sum_{u=1}^{N}x_u + \lambda\theta_0$$
$$\theta = \frac{\sum_{u=1}^{N}x_u + \lambda\theta_0}{N+\lambda}$$

$$\theta = \frac{\sum_{u=1}^{N}x_u + \lambda\theta_0}{N+\lambda} = \theta_{RR}.$$

Exercise 16.

1. $\angle(P_1, P_2, P_3) = \sum_{i=1}^{N} \sum_{j=1}^{3} P(j|x_i) \ln P_j + \lambda \left(\sum_{j=1}^{3} P_j - 1 \right)$

2. $\dfrac{d\angle}{dP_j} = 0 \Rightarrow \dfrac{1}{P_j} \sum_{i=1}^{N} P(j|x_i) = -\lambda = 0$

   $P_j = -\dfrac{\sum_{i=1}^{N} P(j|x_i)}{\lambda}$

3. $\sum_{j=1}^{3} P_j = 1$

   $-\sum_{j=1}^{3} \sum_{i=1}^{N} \dfrac{P(j|x_i)}{\lambda} = 1 \Rightarrow \lambda = -\sum_{j=1}^{3} \sum_{i=1}^{N} P(j|x_i)$

   $\sum_{j=1}^{3} P(j|x_i) = 1 \Rightarrow \lambda = -N.$

4. $\lambda = -N \Rightarrow$ From 2. $P_j = \bullet \dfrac{\sum_{i=1}^{N} P(j|x_i)}{N}$

Exercise 17.

$$\frac{\partial \sum_{i=1}^{N} P(j\mid x_i)\, \ln p(x_i\mid j;\mu_j)}{\partial \mu_j} = 0.$$

$$\frac{\partial \sum_{i=1}^{N} P(j\mid x_i)\, \ln\left[\frac{1}{\left|2\pi\right|^{d}\left|\Sigma_j\right|^{1/2}}\exp\left(-\frac{(x_i-\mu_j)^{T}\Sigma_j^{-1}(x_i-\mu_j)}{2}\right)\right]}{\partial \mu_j}$$

$$\frac{\partial \sum_{i=1}^{N} P(j\mid x_i)\left[\ln\frac{1}{\left|2\pi\right|^{d}\left|\Sigma_j\right|^{1/2}} - \frac{(x_i-\mu_j)^{T}\Sigma_j^{-1}(x_i-\mu_j)}{2}\right]}{\partial \mu_j} = 0$$

$$\frac{\partial \sum_{i=1}^{N} P(j\mid x_i)\, \ln\frac{1}{\left|2\pi\right|^{d}\left|\Sigma_j\right|^{1/2}}}{\partial \mu_j}^{\,0} - \frac{\partial \sum_{i=1}^{N} P(j\mid x_i)(x_i-\mu_j)^{T}\Sigma_j^{-1}(x_i-\mu_j)}{\partial \mu_j}{2} = 0.$$

$$\frac{-1}{2}\, \frac{\partial \sum_{i=1}^{N} P(j\mid x_i)(x_i-\mu_j)^{T}\Sigma_j^{-1}(x_i-\mu_j)}{\partial \mu_j} = 0.$$

$$\sum_{i=1}^{N} P(j\mid x_i)\left(-2(x_i-\mu_j)\Sigma_j^{-1}\right) = 0$$

$$\sum_{i=1}^{N} P(j\mid x_i)(x_i-\mu_j) = 0 \Rightarrow$$

$$\sum_{i=1}^{N} P(j\mid x_i)\, x_i = \mu_j \sum_{i=1}^{N} P(j\mid x_i) \Rightarrow$$

$$\boxed{\mu_j = \frac{\sum_{i=1}^{N} P(j\mid x_i)\, x_i}{\sum_{i=1}^{N} P(j\mid x_i)}}$$

```
In [4]:  import scipy.io as sio
         import numpy as np
         import matplotlib.pyplot as plt

         Dataset = sio.loadmat('Dataset.mat')
         X1 = Dataset['X1']
         X2 = Dataset['X2']
```

```
In [5]:  import scipy.io as sio
         import numpy as np
         from sklearn.mixture import GaussianMixture
         import matplotlib.pyplot as plt
         from mpl_toolkits.mplot3d import Axes3D

         # Load the dataset
         Dataset = sio.loadmat('Dataset.mat')
         X1 = Dataset['X1']
         X2 = Dataset['X2']

         # Plotting just the X1 and X2 points initially
         fig = plt.figure(figsize=(12, 6))

         # Plot for X1
         ax = fig.add_subplot(121, projection='3d')
         ax.scatter(X1[:, 0], X1[:, 1], X1[:, 2], c='blue', alpha=0.6, label='X1 Data')
         ax.set_title('X1 Data')
         ax.set_xlabel('X')
         ax.set_ylabel('Y')
         ax.set_zlabel('Z')
         ax.legend()

         # Plot for X2
         ax = fig.add_subplot(122, projection='3d')
         ax.scatter(X2[:, 0], X2[:, 1], X2[:, 2], c='green', alpha=0.6, label='X2 Data')
         ax.set_title('X2 Data')
         ax.set_xlabel('X')
         ax.set_ylabel('Y')
         ax.set_zlabel('Z')
         ax.legend()

         plt.tight_layout()
         plt.show()
```
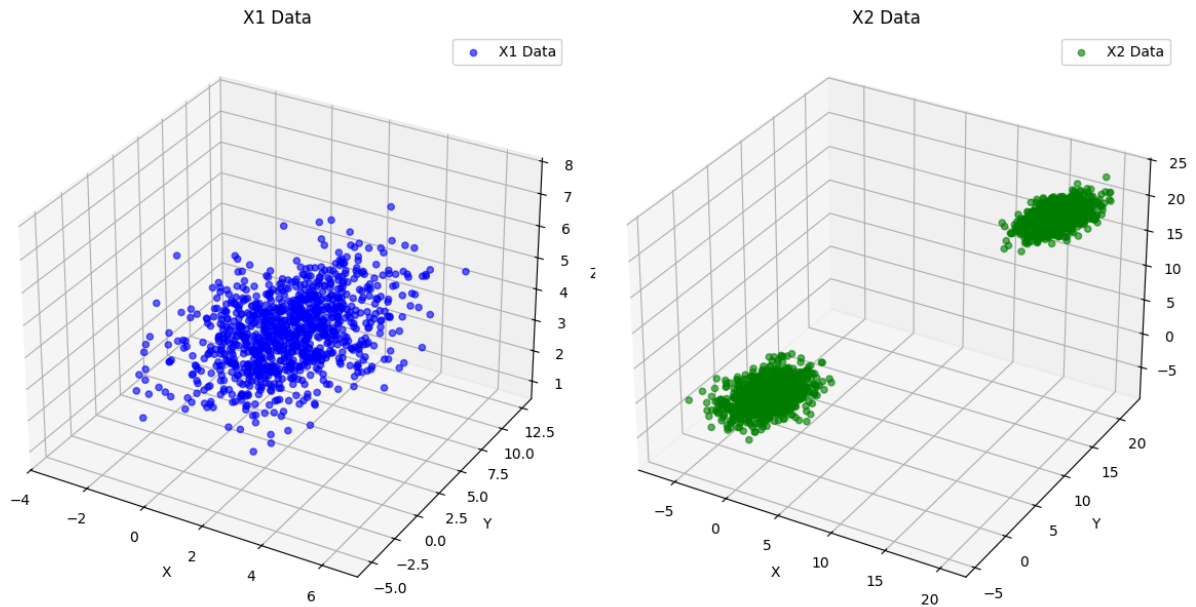
X1 Data   X2 Data

```
In [10]:  import scipy.io as sio
          import numpy as np
          from sklearn.mixture import GaussianMixture
          import matplotlib.pyplot as plt
          from mpl_toolkits.mplot3d import Axes3D

          # Load the dataset
          Dataset = sio.loadmat('Dataset.mat')
          X1 = Dataset['X1']
          X2 = Dataset['X2']

          # Define the points for which we want to estimate the pdf
          points_X1 = np.array([[2.01, 2.99, 3.98, 5.02],
                                [20.78, -15.26, 19.38, -25.02],
                                [3.08, 3.88, 4.15, 6.02]])

          points_X2 = np.array([[0.05, 0.15, -0.12, -0.08],
                                [7.18, 7.98, 9.12, 9.94],
                                [3.48, 4.01, 4.55, 4.96],
                                [20.78, -15.26, 19.38, -25.02]])

          # Compute the mean for X1 and X2
          mean_X1 = np.mean(X1, axis=0)
          mean_X2 = np.mean(X2, axis=0)

          # Fit GMMs to X1 and X2
          gmm_X1 = GaussianMixture(n_components=1, covariance_type='full').fit(X1)
          pdf_values_X1 = np.exp(gmm_X1.score_samples(points_X1))

          gmm_X2 = GaussianMixture(n_components=1, covariance_type='full').fit(X2)
          pdf_values_X2 = np.exp(gmm_X2.score_samples(points_X2))

          # Plotting the datasets with pdf values visualized
          fig = plt.figure(figsize=(12, 6))

          # Plot for X1
```

```
ax = fig.add_subplot(121, projection='3d')
sc = ax.scatter(X1[:, 0], X1[:, 1], X1[:, 2], c='blue', alpha=0.6, label='X1 Data')
ax.scatter(points_X1[:, 0], points_X1[:, 1], points_X1[:, 2], c=pdf_values_X1, cmap
ax.scatter(mean_X1[0], mean_X1[1], mean_X1[2], c='red', s=100, label='Mean of X1',
plt.colorbar(sc, ax=ax, label='X1 Data')
ax.set_title('X1 Data with Distance from Mean and PDF Values')
ax.set_xlabel('X')
ax.set_ylabel('Y')
ax.set_zlabel('Z')
ax.legend()

# Plot for X2
ax = fig.add_subplot(122, projection='3d')
sc = ax.scatter(X2[:, 0], X2[:, 1], X2[:, 2], c='green', alpha=0.6, label='X2 Data'
ax.scatter(points_X2[:, 0], points_X2[:, 1], points_X2[:, 2], c=pdf_values_X2, cmap
ax.scatter(mean_X2[0], mean_X2[1], mean_X2[2], c='red', s=100, label='Mean of X2',
plt.colorbar(sc, ax=ax, label='X2 Data')
ax.set_title('X2 Data with Distance from Mean and PDF Values')
ax.set_xlabel('X')
ax.set_ylabel('Y')
ax.set_zlabel('Z')
ax.legend()

plt.tight_layout()
plt.show()

# Print the estimated pdf values for the points
print("Estimated pdf values for X1 points:", pdf_values_X1)
print("Estimated pdf values for X2 points:", pdf_values_X2)
```
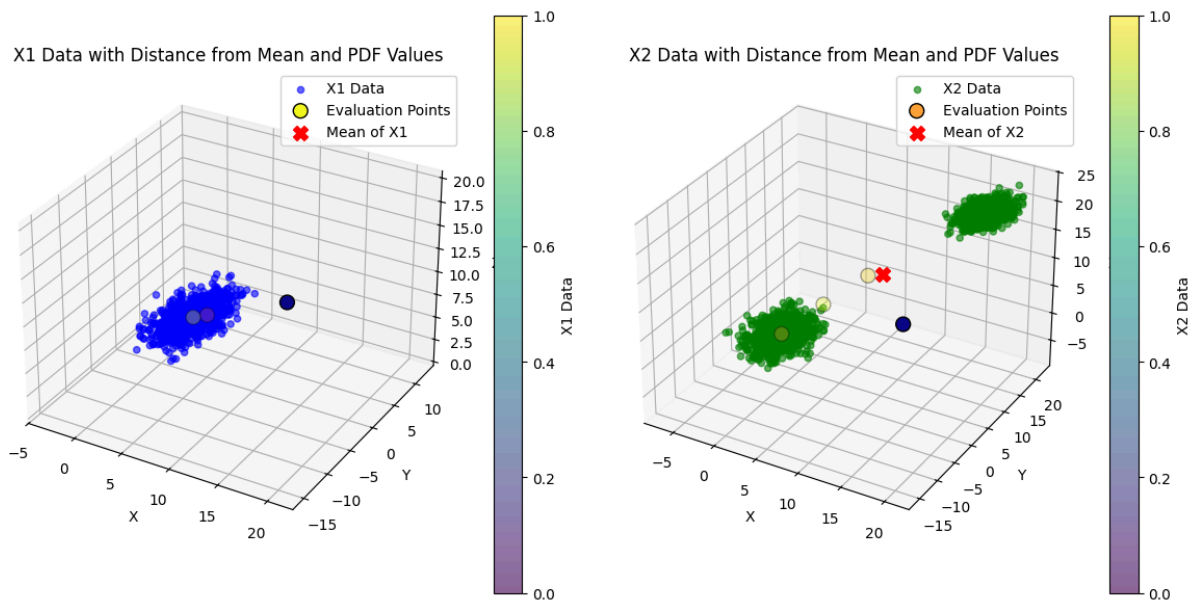


```
Estimated pdf values for X1 points: [3.37903694e-003 1.28655830e-206 1.95207408e-00
3]
Estimated pdf values for X2 points: [3.02745641e-004 3.57413507e-004 3.85782183e-004
1.53985535e-131]
```

Exercise 18 Solution: PDF Estimation for Data Points in X1 and X2 Using Gaussian Mixture Modeling

This code uses a Gaussian Mixture Model (GMM) to estimate the probability density function (pdf) values at specific points for two datasets, `X1` and `X2`. We also enrich the pdf modeling by calculating the distance of each data vector from the mean of its dataset, which helps visualize the data spread and density around the mean. Here's a breakdown of the approach:

Load Dataset and Define Points for PDF Estimation

We start by loading the datasets `X1` and `X2` from `Dataset.mat`. We also define a set of points for which we want to estimate the pdf values based on each dataset.

- **Dataset**: Each dataset (`X1` and `X2`) contains 4-dimensional data vectors.
- **Points**: We specify points (e.g., `(2.01, 2.99, 3.98, 5.02)`) at which we aim to estimate pdf values using a Gaussian Mixture Model.

Compute Mean and Distance from Mean

For both `X1` and `X2`, we calculate:

- The **mean vector**, which provides the center of the data distribution.
- The **Euclidean distance** of each data vector from this mean vector.

These distances give us insight into how spread out each dataset is around its mean. In later steps, we use these distances to visualize the density and structure of the data in 3D plots.

Fit Gaussian Mixture Models (GMM)

We fit a Gaussian Mixture Model (GMM) with one component to each dataset (`X1` and `X2`). This model estimates a Gaussian distribution for each dataset, capturing its pdf structure. Using the `score_samples` method, we compute the log probability density for the specified points, which we exponentiate to get the pdf values.

- **PDF Estimation**: The GMM provides a parametric approach to estimating the pdf, using the underlying Gaussian structure of the data.
- **Output**: We obtain pdf values for each specified point in `X1` and `X2`.

Visualize Data and PDF Estimation with 3D Scatter Plots

We create two 3D scatter plots (one for each dataset) to visualize the distribution of data points and estimated pdf values. Here's how each plot is structured:

Plot Structure

- **Dataset Points**: The data points in `X1` and `X2` are shown with color intensity representing their **distance from the mean**. Points closer to the mean have a lighter color, while those farther away have a darker color.
- **Evaluation Points**: The specified points for pdf estimation are marked with black edges and sized slightly larger for visibility. The color of these points represents the **pdf value** estimated by the GMM model.
- **Color Bar**: Each subplot includes a color bar that indicates the distance from the mean, enhancing our understanding of the data spread relative to the center.

Interpretation of Results

- **PDF Values**: The estimated pdf values are displayed for each evaluation point, giving an understanding of how likely these points are according to the Gaussian distribution fitted to `X1` and `X2`.
- **Data Structure and Spread**: The color gradient from each plot reflects the distances from the mean, helping us observe how data points are distributed in relation to the center.