Machine Learning Programming

GMM Overview

Lecture : Prof. Aude Billard (aude.billard@epfl.ch)

Teaching Assistants :
Nadia Figueroa, Laila El Hamamsy, Hala Khodr and Lukas Huber

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

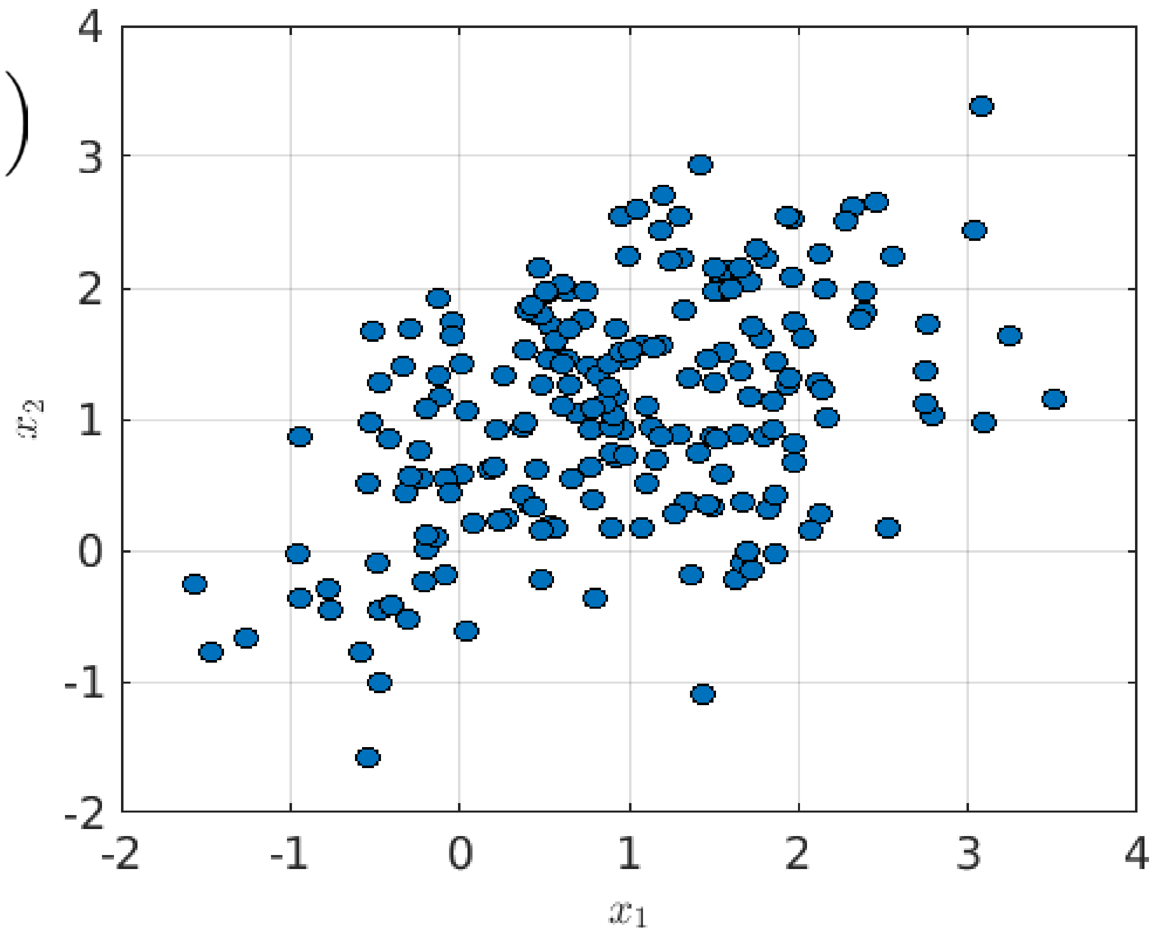
Data modeling with Gaussian probability density function

$$p(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(\mathbf{x}-\mu)^T(\Sigma)^{-1}(\mathbf{x}-\mu)\right)}$$

$\mathbf{x} \in \mathbb{R}^N$: datapoints

$\mu \in \mathbb{R}^N$: mean vector

$\Sigma \in \mathbb{R}^{N\times N}$: Covariance Matrix



Data modeling with Gaussian probability density function

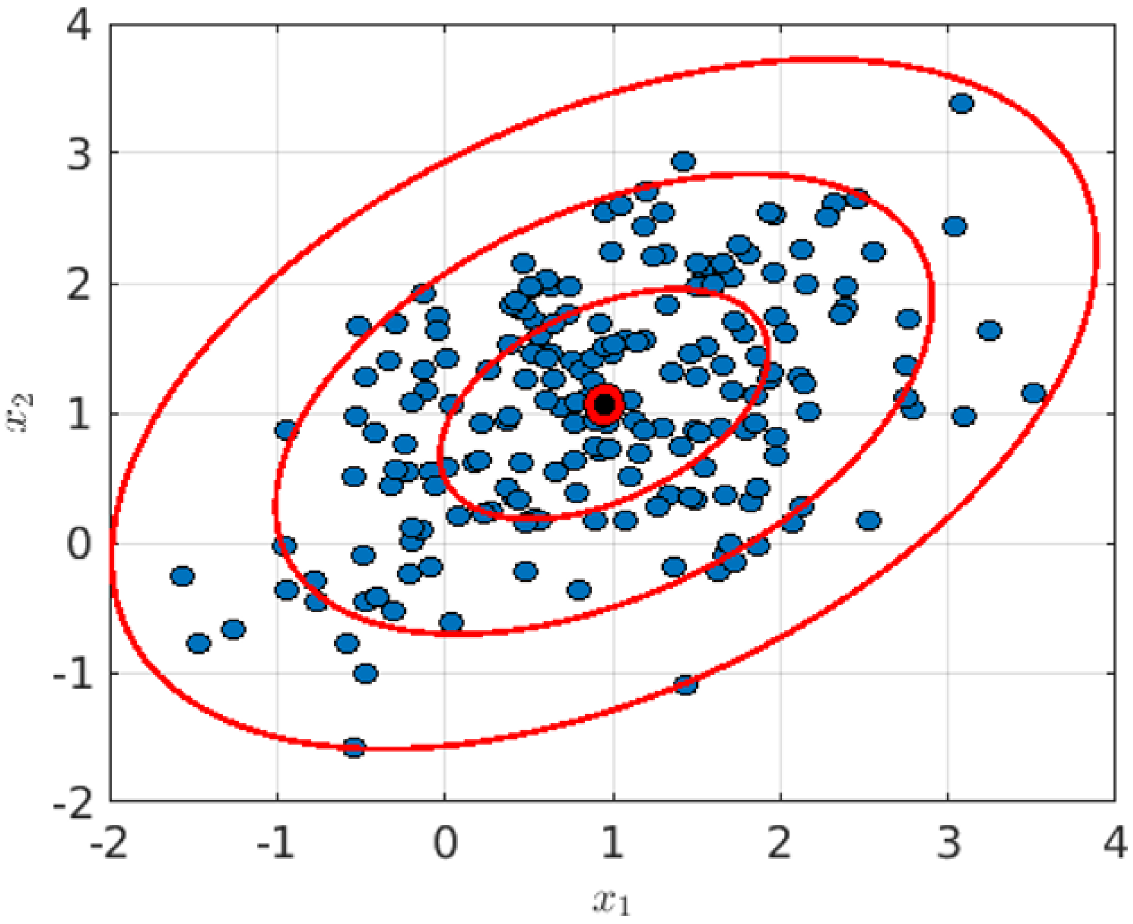
$$p(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \boxed{(\Sigma)}^{-1}(\mathbf{x}-\boxed{\mu})\right)}$$

$$\mathbf{x} \in \mathbb{R}^N \quad \text{: datapoints}$$
$$\mu \in \mathbb{R}^N \quad \text{: mean vector}$$
$$\Sigma \in \mathbb{R}^{N\times N} \quad \text{: Covariance Matrix}$$

Estimate parameters through **Maximum Likelihood** Optimization

$$\max_{\mu,\Sigma} \quad \log \mathcal{L}(\mu,\Sigma|\mathbf{X}) \quad = \max_{\mu,\Sigma} \quad \log p(\mathbf{X}|\mu,\Sigma)$$

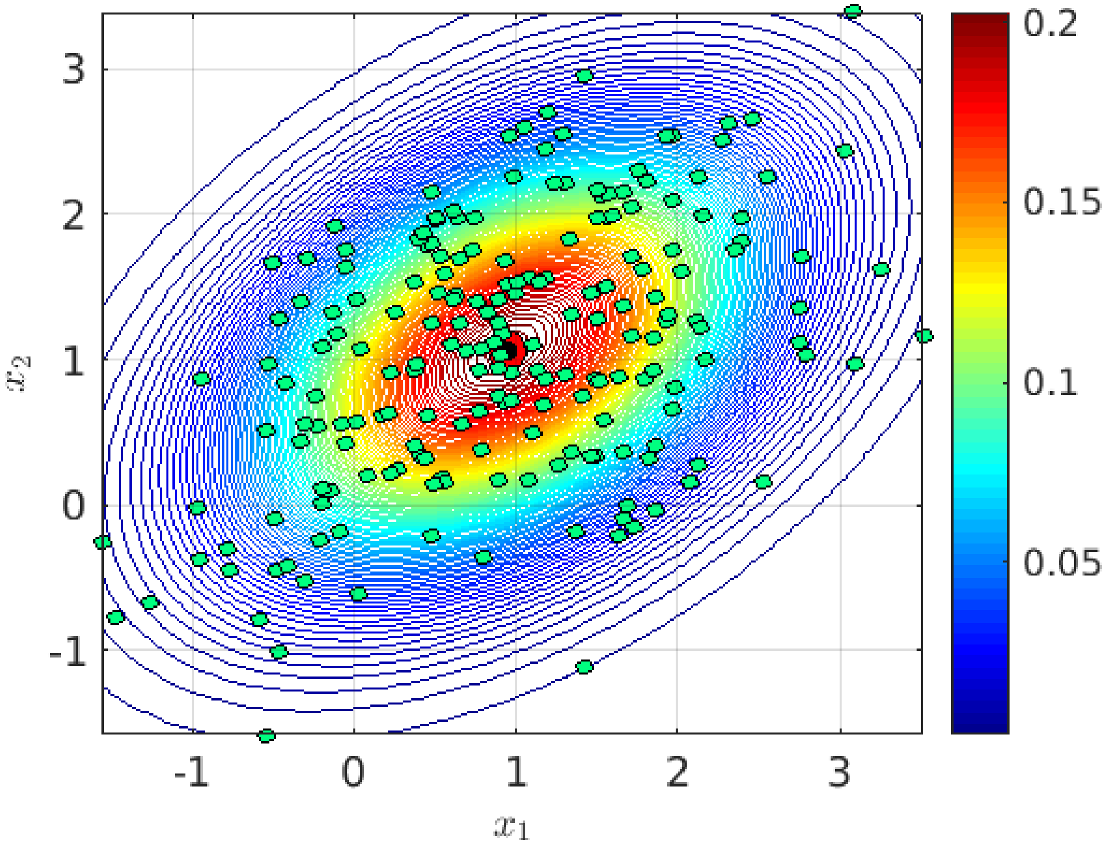
$$\frac{\partial}{\partial\mu} p(\mathbf{X}|\mu,\Sigma) = 0 \quad \text{and} \quad \frac{\partial}{\partial\Sigma} p(\mathbf{X}|\mu,\Sigma) = 0$$



Data modeling with Gaussian probability density function

$$p(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(\mathbf{x}-\mu)^T(\Sigma)^{-1}(\mathbf{x}-\mu)\right)}$$

$$\begin{aligned} \mathbf{x} &\in \mathbb{R}^N && : \text{datapoints} \\ \mu &\in \mathbb{R}^N && : \text{mean vector} \\ \Sigma &\in \mathbb{R}^{N\times N} && : \text{Covariance Matrix} \end{aligned}$$



Estimate parameters through **Maximum Likelihood** Optimization

$$\max_{\mu,\Sigma} \log \mathcal{L}(\mu,\Sigma|\mathbf{X}) = \max_{\mu,\Sigma} \log \boxed{p(\mathbf{X}|\mu,\Sigma)}$$

$$\frac{\partial}{\partial\mu} p(\mathbf{X}|\mu,\Sigma) = 0 \quad \text{and} \quad \frac{\partial}{\partial\Sigma} p(\mathbf{X}|\mu,\Sigma) = 0$$

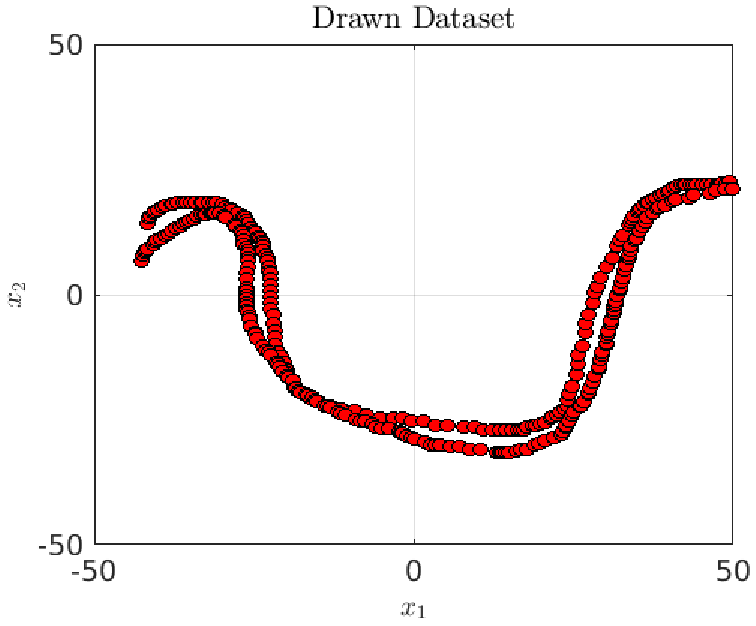
Closed form solution: straight-forward to compute.

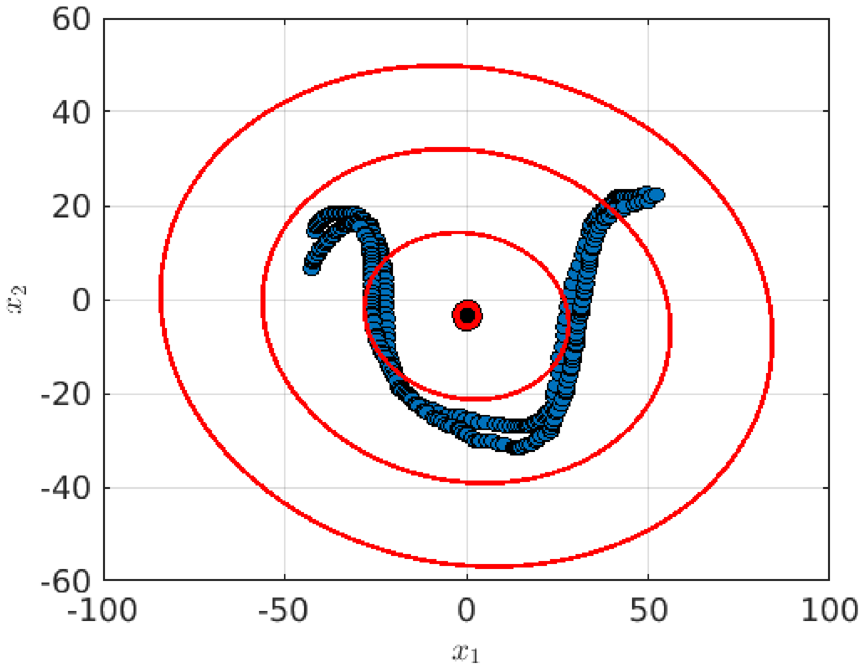
Likelihood function

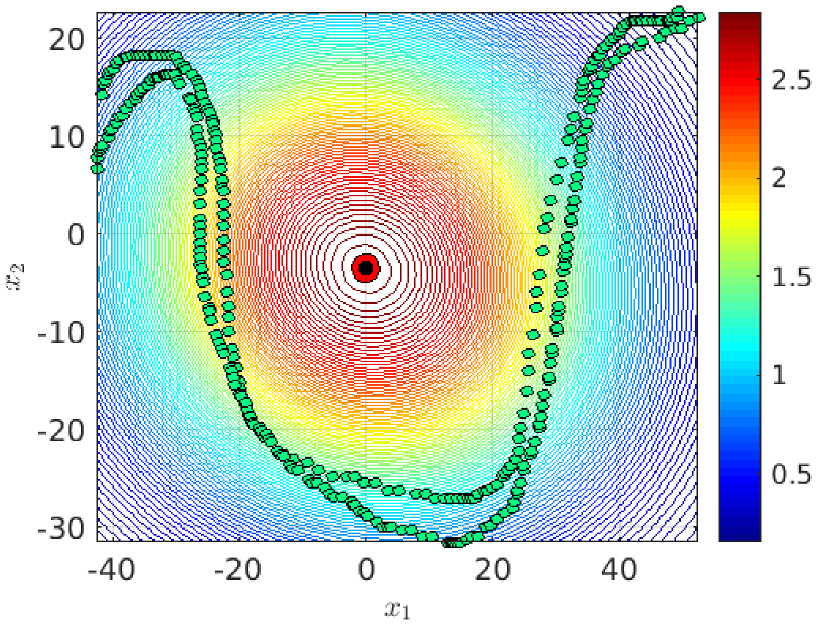
$$p(\mathbf{X}|\mu,\Sigma) = \prod_{i=1}^{M} p(\mathbf{x}^i|\mu,\Sigma)$$

What if we have more 'complex' data?

$$p(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(\mathbf{x}-\mu)^T(\Sigma)^{-1}(\mathbf{x}-\mu)\right)}$$







Not a very good fit! :(

Data modeling with Gaussian Mixture Models

$$p(\mathbf{x}|\Theta) = \sum_{k=1}^{K} \alpha_k \boxed{p(\mathbf{x}|\mu^k, \mathbf{\Sigma}^k)}$$

k-th Gaussian function (task 1)

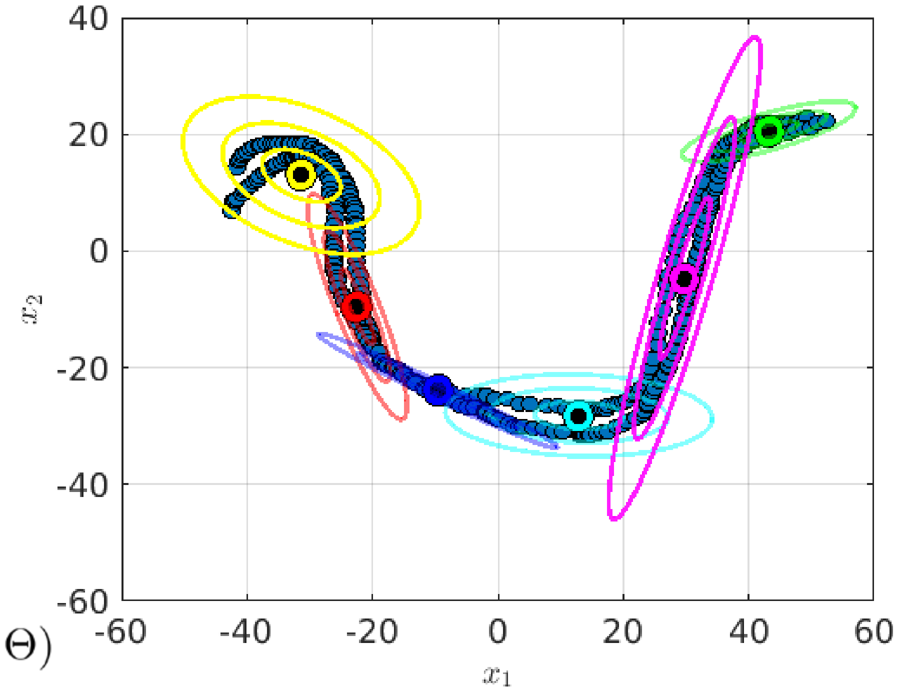
$$p(\mathbf{x}|\mu^k, \Sigma^k) = \frac{1}{(2\pi)^{N/2}|\Sigma^k|^{1/2}} e^{(-\frac{1}{2}(\mathbf{x}-\mu^k)^T(\Sigma^k)^{-1}(\mathbf{x}-\mu^k))}$$

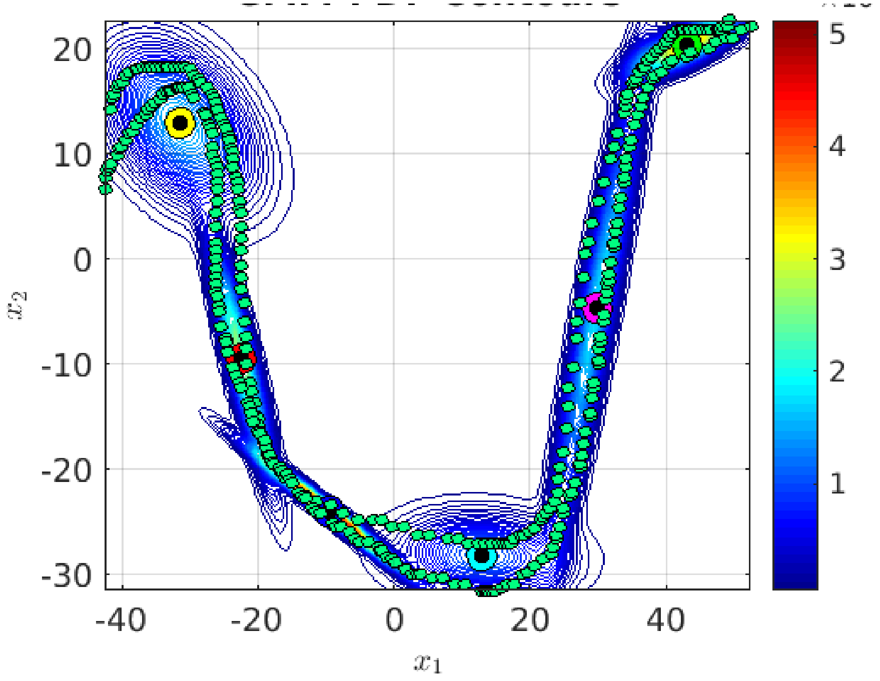
GMM parameters

$$\Theta = \{\theta^1, \ldots, \theta^k\}$$

$$\theta_k = \{\alpha_k, \mu^k, \Sigma^k\}$$

$$\sum_{k=1}^{K} \alpha_k = 1$$





Estimate parameters through **Maximum Likelihood** Optimization

$$\max_{\Theta} \quad \log \mathcal{L}(\Theta|\mathbf{X}) = \max_{\Theta} \quad \log p(\mathbf{X}|\Theta)$$

$$\max_{\Theta} \quad \log \mathcal{L}(\Theta|\mathbf{X}) = \max_{\Theta} \quad \log \prod_{i=1}^{M} \sum_{k=1}^{K} \alpha_k p(\mathbf{x}^i|\mu^k, \mathbf{\Sigma}^k)$$

GMM Likelihood function **(task 2)**

NO closed form solution! Solved through EXPECTATION-MAXIMIZATION Algorithm **(task 4, 5)**

Expectation Maximization (EM) for GMM

1. **Initialization step:** Initialize priors $\alpha = \{\alpha_1, \ldots, \alpha_k\}$, means $\mu = \{\mu^1, \ldots, \mu^k\}$ and Covariance matrices $\mathbf{\Sigma} = \{\mathbf{\Sigma}^1, \ldots, \mathbf{\Sigma}^K\}$.
2. **Expectation Step:** For each Gaussian $k \in \{1, \ldots, K\}$, compute the probability that it is responsible for each point $\mathbf{x}^i$ in the dataset.
3. **Maximization Step:** Re-estimate the priors $\alpha = \{\alpha_1, \ldots, \alpha_K\}$, means $\mu = \{\mu^1, \ldots, \mu^K\}$ and Covariance matrices $\mathbf{\Sigma} = \{\mathbf{\Sigma}^1, \ldots, \mathbf{\Sigma}^K\}$
4. Go back to step 2 and repeat until the $\log \mathcal{L}(\Theta|\mathbf{X})$ stabilizes.

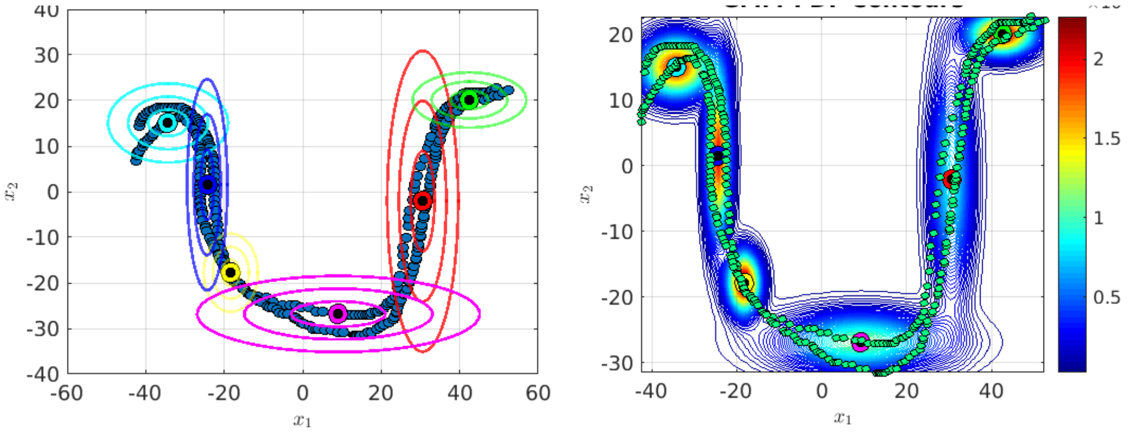
Expectation Maximization (EM) for GMM

Initialize means with K-means (task 4) and Covariance matrices (task3)

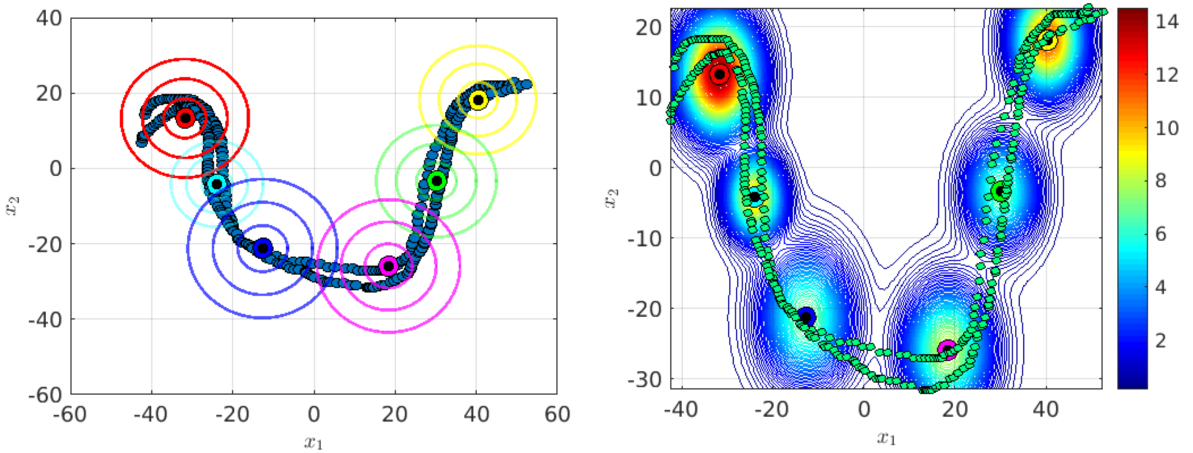
1. **Initialization step:** Initialize priors $\alpha = \{\alpha_1, \ldots, \alpha_k\}$, means $\mu = \{\mu^1, \ldots, \mu^k\}$ and Covariance matrices $\mathbf{\Sigma} = \{\mathbf{\Sigma}^1, \ldots, \mathbf{\Sigma}^K\}$.
2. **Expectation Step:** For each Gaussian $k \in \{1, \ldots, K\}$, compute the probability that it is responsible for each point $\mathbf{x}^i$ in the dataset.
3. **Maximization Step:** Re-estimate the priors $\alpha = \{\alpha_1, \ldots, \alpha_K\}$, means $\mu = \{\mu^1, \ldots, \mu^K\}$ and Covariance matrices $\mathbf{\Sigma} = \{\mathbf{\Sigma}^1, \ldots, \mathbf{\Sigma}^K\}$
4. Go back to step 2 and repeat until the $\log \mathcal{L}(\Theta|\mathbf{X})$ stabilizes.

Different types of Covariance matrices (task3)

Diagonal Covariance matrices



Isotropic Covariance matrices



Expectation Maximization (EM) for GMM

1. **Initialization step:** Initialize priors $\alpha = \{\alpha_1, \ldots, \alpha_k\}$, means $\mu = \{\mu^1, \ldots, \mu^k\}$ and Covariance matrices $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K\}$.
2. **Expectation Step:** For each Gaussian $k \in \{1, \ldots, K\}$, compute the probability that it is responsible for each point $\mathbf{x}^i$ in the dataset.
3. **Maximization Step:** Re-estimate the priors $\alpha = \{\alpha_1, \ldots, \alpha_K\}$, means $\mu = \{\mu^1, \ldots, \mu^K\}$ and Covariance matrices $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K\}$
4. Go back to step 2 and repeat until the $\log \mathcal{L}(\Theta|\mathbf{X})$ stabilizes.

EM iterative procedure following Eq. 8-12 in assignment description (task5)

(task 6) and (task 7) Implement fitting GMMs through Model Selection approach (similar to K-Means)