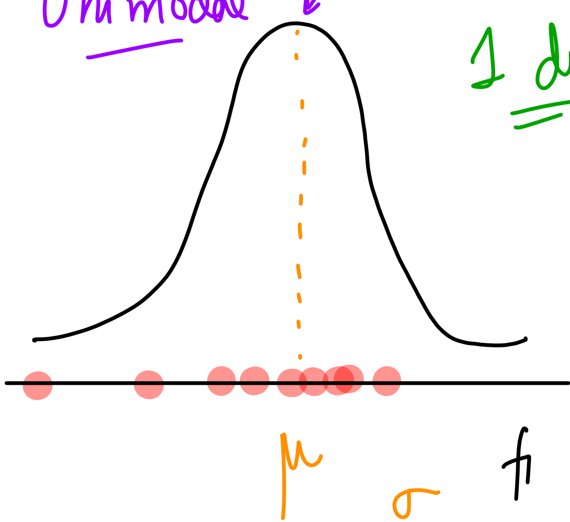
GMM

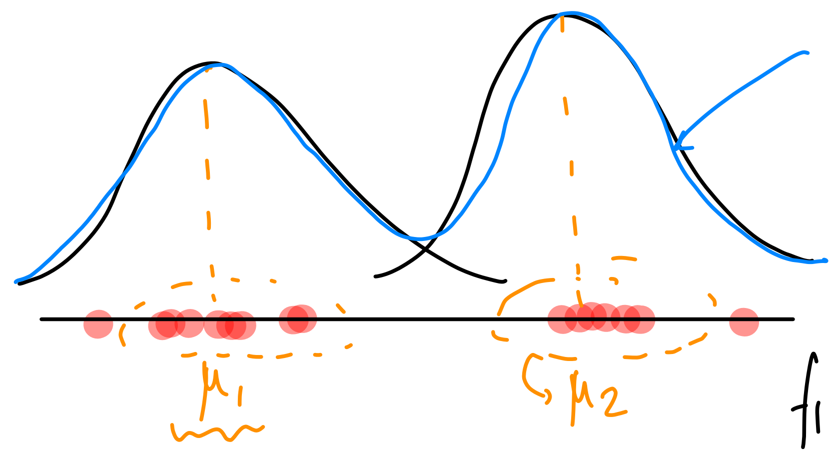
Gaussian Mixture Models

Unimodal

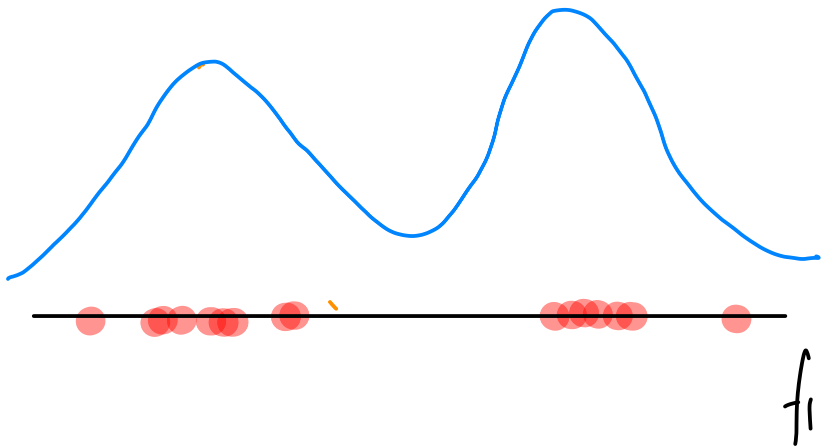
1 dim



multimodal



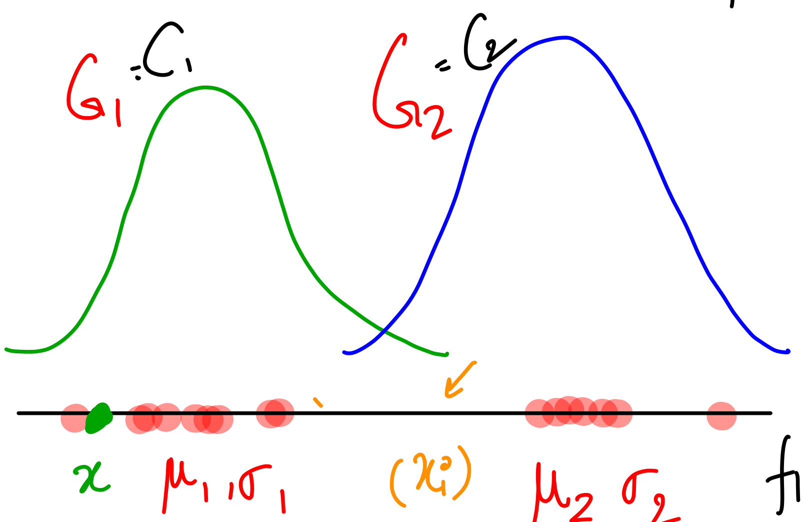
Mixture of Gaussian.



PDF

$x \to G_1 = 0.2$

$G_2 = 0.0001$

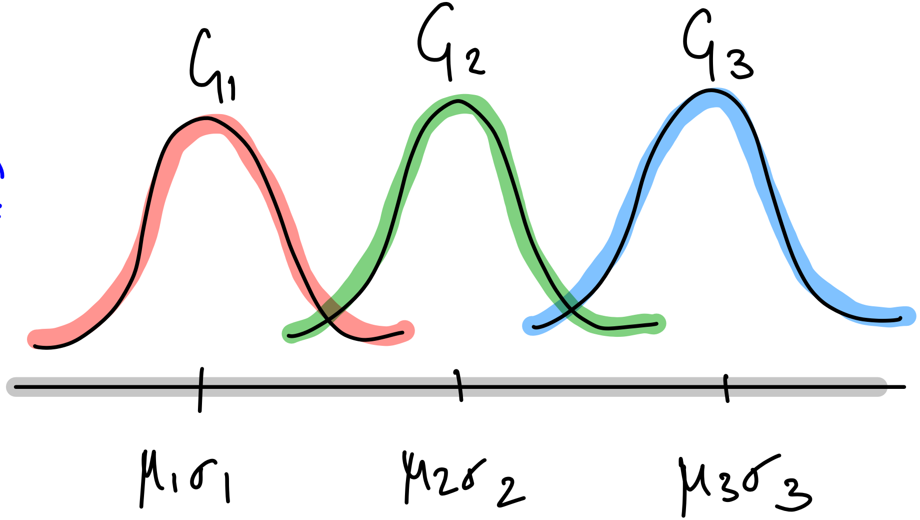


$x_i \to G_1 \to 0.1$

$\to G_2 \to 0.9$

Gaussian Mixture Model

1D GMM



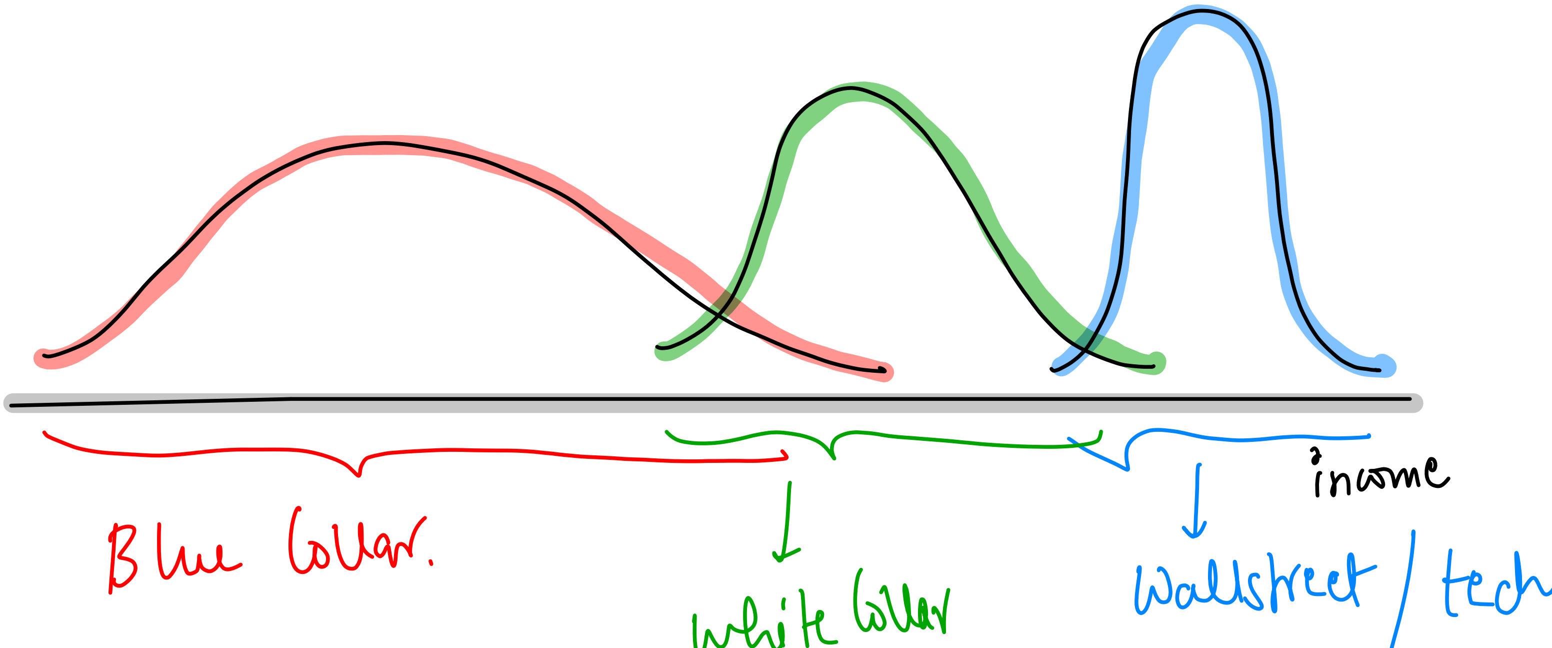
$x_i \rightarrow G_1 : P(y_i = 1)$

$G_2 : P(y_i = 2)$

$G_3 : P(y_i = 3)$

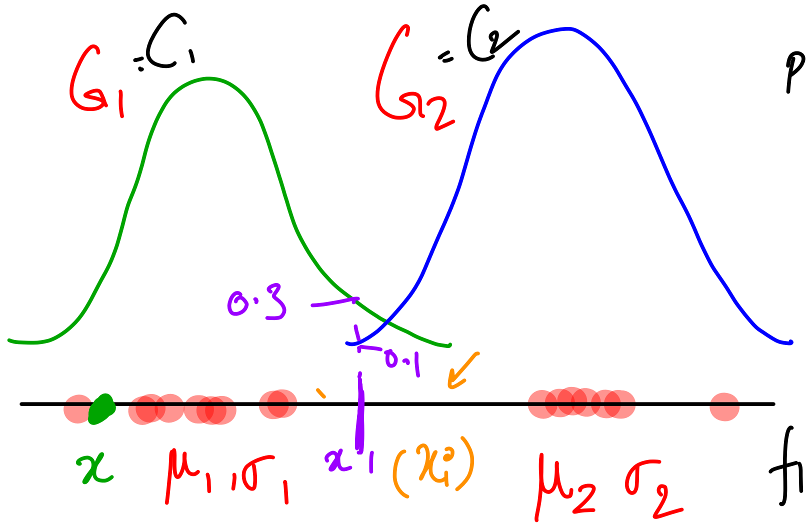
ecommerce $\rightarrow$ price conscious 40%

$\rightarrow$ wealthy 60%



income
Blue Collar.
white Collar
Wallstreet / tech

Soft Assignment



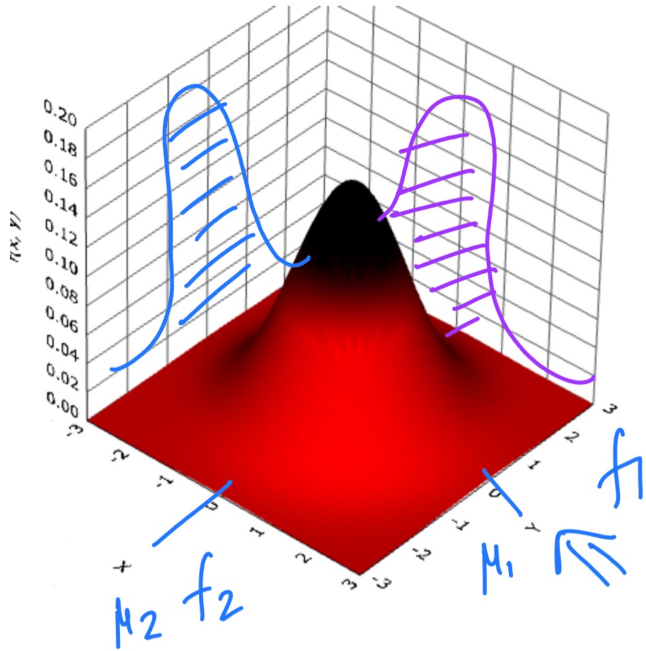
$$\text{PDF} = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

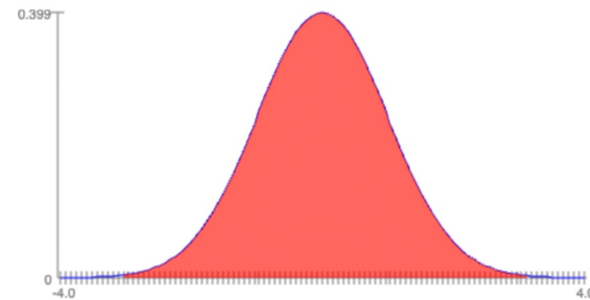
$$x_1 \rightarrow G_1 = 0.3$$
$$G_2 = 0.1$$

Soft assignment

KNN

$$\left.\begin{aligned} x_i &\rightarrow C_1 \\ &\rightarrow C_2 \end{aligned}\right\}$$ Hard Assignment.


0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
x
y
f_1
μ_1
μ_2 f_2

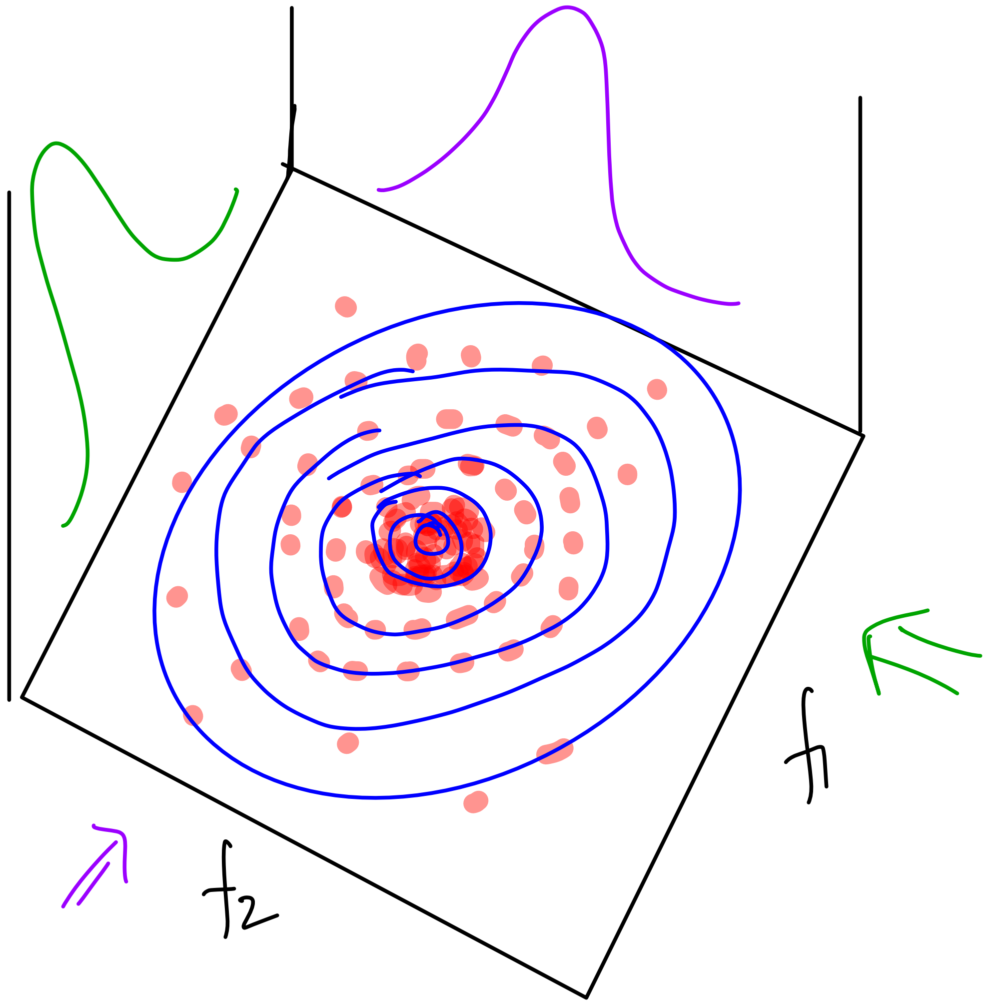

0.399
0
-4.0
4.0

2 dim
data

hill
3d

d dim

hill → d+1
dim



Contour

2D Gaussian
⇓
3D space.

1D . $N(\mu, \sigma)$

$\mu \in \mathbb{R}^1$

$\sigma \in \mathbb{R}^1$

2D . $N_2(\mu, \Sigma)$

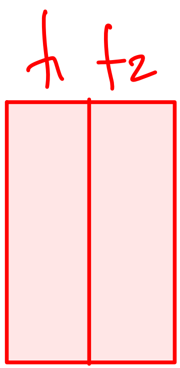
$\mu \in \mathbb{R}^2$

Covariance Matrix

Σ: 2×2

$[\mu_1, \mu_2]$

$$\begin{array}{c|cc} & f_1 & f_2 \\ \hline f_1 & Cov(f_1 f_1) & Cov(f_1 f_2) \\ f_2 & Cov(f_2 f_1) & Cov(f_2 f_2) \end{array}$$

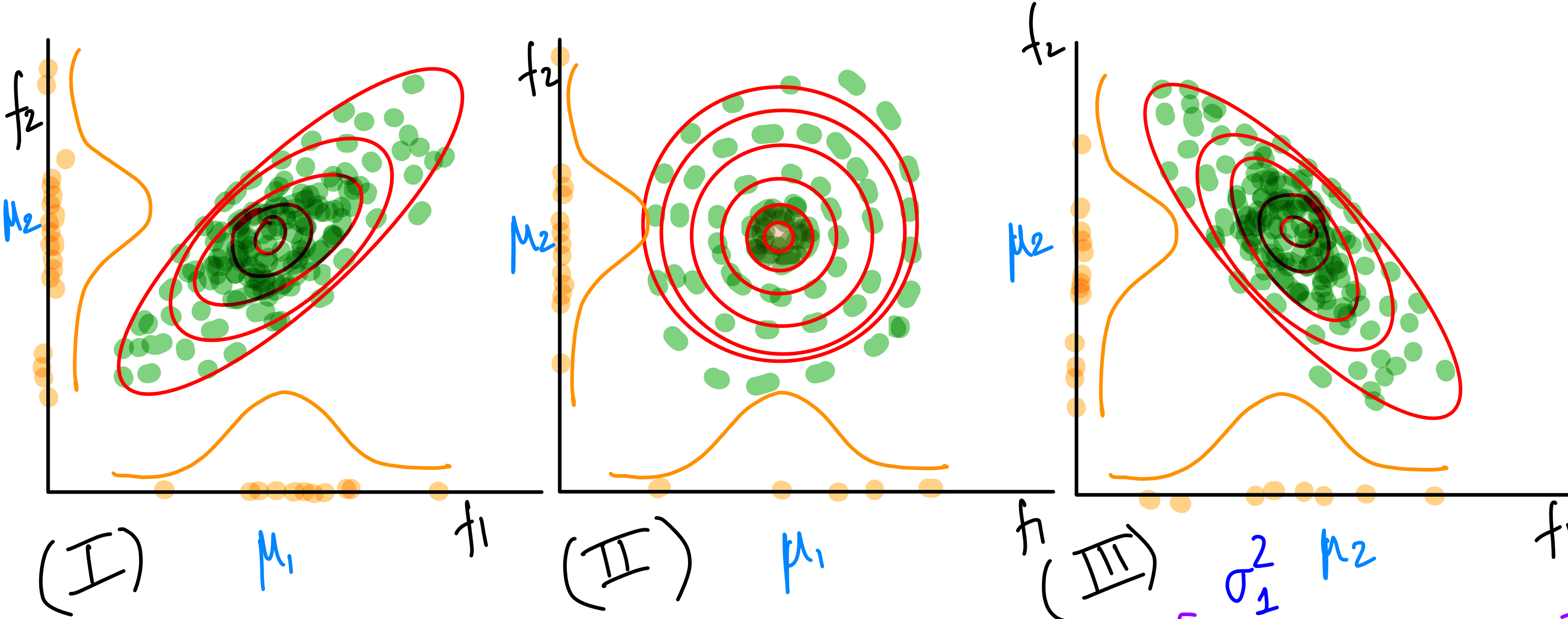
$D =$ 

d-Dim

$N_d(\mu, \Sigma)$

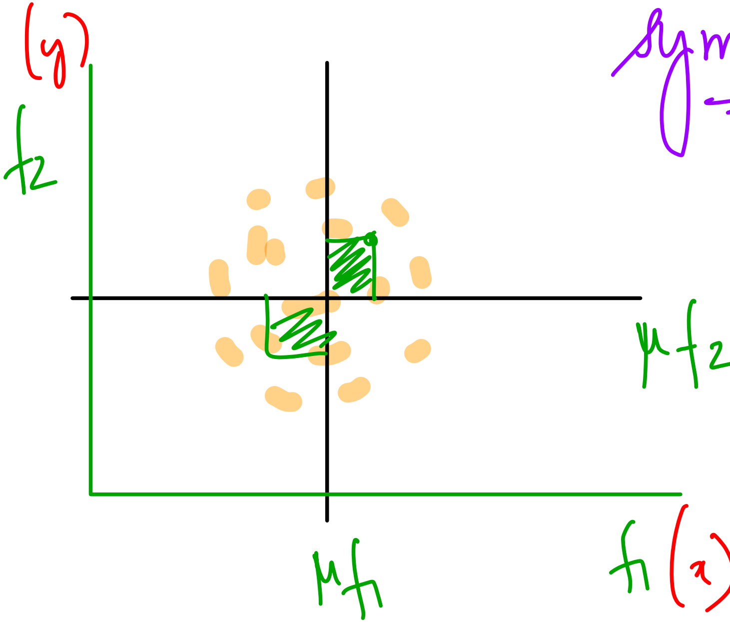
$\mu \in \mathbb{R}^d$

Σ: $d \times d$



$\mu_1, \mu_2, \sigma_1, \sigma_2$

$$\Sigma = \begin{bmatrix} \overset{\sigma_1^2}{Cov(f_1, f_1)} & Cov(f_1, f_2) \\ Cov(f_2, f_1) & \underset{\sigma_2^2}{Cov(f_2, f_2)} \end{bmatrix}$$



Symmetric

$$\Sigma = \begin{bmatrix} \sigma_1^2 & Cov(f_1, f_2) \\ Cov(f_2, f_1) & \sigma_2^2 \end{bmatrix}$$

$Cov(f_1, f_2) \Rightarrow +ve \Rightarrow 1$

$\Rightarrow -ve \Rightarrow 3$

$\Rightarrow 0 \Rightarrow 2$

$$Cov(f_1, f_2) = \frac{1}{n} \sum (x - \bar{x})(y - \bar{y})$$

PDF :

$$(2\pi)^{-k/2}\det(\boldsymbol\Sigma)^{-1/2}\,\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol\mu)^{\mathrm T}\boldsymbol\Sigma^{-1}(\mathbf{x}-\boldsymbol\mu)\right),$$

Inverse of matrix

vector

vector

$A\,B\,B^{-1} = C B^{-1}$

$A = C B^{-1}$

$AB = C$

$Ay = B$

$\frac{Ay}{y} = \frac{B}{y}$

$\boxed{A = B y^{-1}}$

d

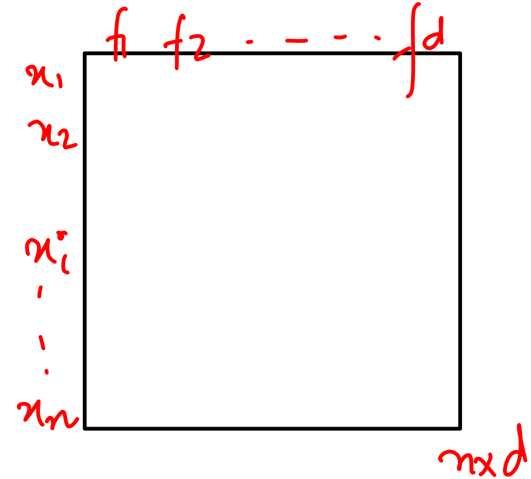
$$A = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$\det A = |A| = A \times D - B \times C$

$$Z = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix} = A(EI - FH) - B(DI - GF) + C(DH - EG)$$

$\det(Z)$

d dimensional data



3 - Gaussian

Find

$$\left.\begin{array}{l} G_1 : \underbrace{\mu_1}_{\in \mathbb{R}^d},\ \underbrace{\Sigma_1}_{\mathbb{R}^{d\times d}} \\ G_2 : \mu_2,\ \Sigma_2 \\ G_3 .\ \mu_3,\ \Sigma_3 \end{array}\right\} \underline{\text{params}}$$

$x_1 \to G_1 .\ p_{i1}$

$x_1 \to G_2 : p_{i2}$

$x_1 \to G_3 : p_{i3}$

$\overline{\quad 1 \quad}$

1 D

PDF of 1 D Gaussian

PDF	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$$x_i^i = P(x_i \mid G_j)$$

$$P(x_i \mid G_1) = 0.2 / 0.8 =$$

$$P(x_i \mid G_2) = 0.1 / 0.8 =$$

$$P(x_i \mid G_3) = 0.5 / 0.8 =$$

$$\underline{\overline{1}}$$

recap

→ 1D GMM

→ d-dim Gaussian-dis $(\mu_d, \Sigma_{d \times d})$

→ d-dim GMM

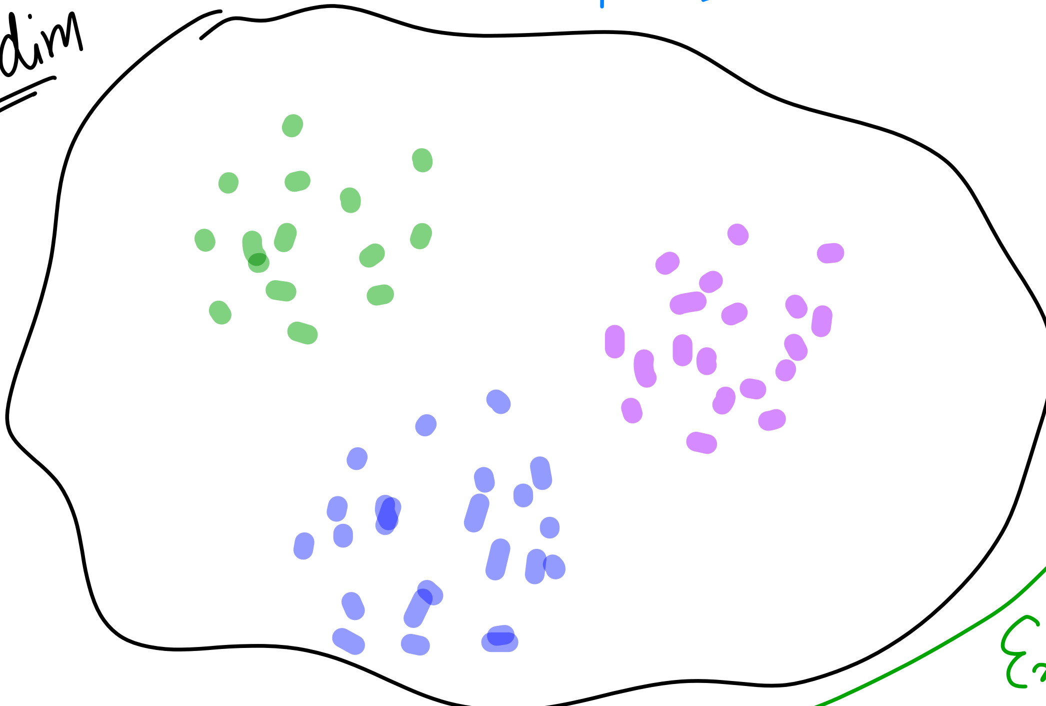
→ x_i → $P(x_i | G_j)$

PDF

$\left\{ \begin{array}{l} \text{1 dim PDF function} \\ \text{d dim " "} \end{array} \right\}$

d dim

probabilities



Find K Gaussian. given $\mathcal{D} \in \mathbb{R}^d$

μ_i , $\Sigma_{d \times d}$

$\mu_i \rightarrow$ d-dim

$\forall\, i : 1 \rightarrow k$

Expectation

optimisation → EM (Coordinate Ascent)

GD.

maximisation

$$X = x_1, x_2, x_3, x_4 \ldots x_n$$

1D - Gaussian $\qquad X \sim N(\mu, \sigma)$

dD - Gaussian $\qquad X \sim N_d\big(\overbrace{\mu_d, \underbrace{\Sigma_{d \times d}}_{\text{symmetric}}}^{\text{params.}}\big)$

$$D = \begin{array}{c} f_1 \; f_2 \; \ldots \; f_d \\ \begin{bmatrix} \leftarrow x_1 \rightarrow \\ \leftarrow x_2 \rightarrow \\ \vdots \\ \leftarrow x_n \rightarrow \end{bmatrix} \end{array}$$

$$\hat{\mu}_d = \frac{1}{n} \sum_{i=1}^{n} x_i$$

x_i → d dimensional vector

$\hat{\mu}_d$ → d dim vector.

Generative Methods

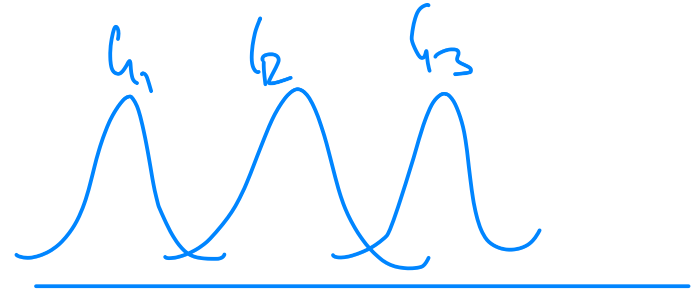
A $P(y = j) \quad \forall i : 1 \to k$

↳ cluster index

$P(y=1) = 0.3$

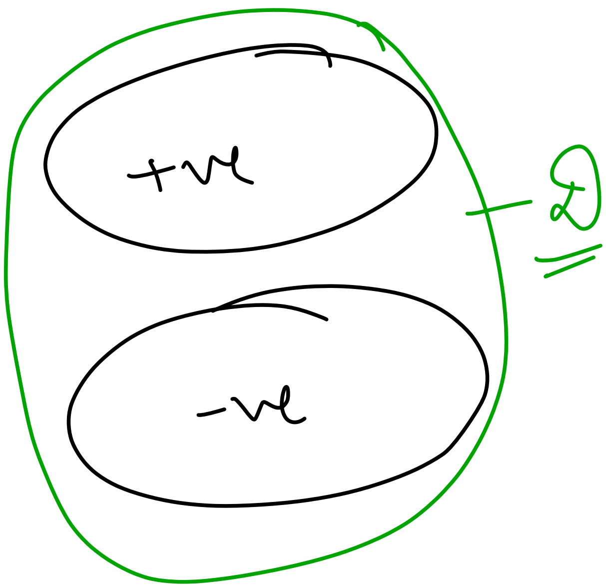
$P(y=2) = 0.4$

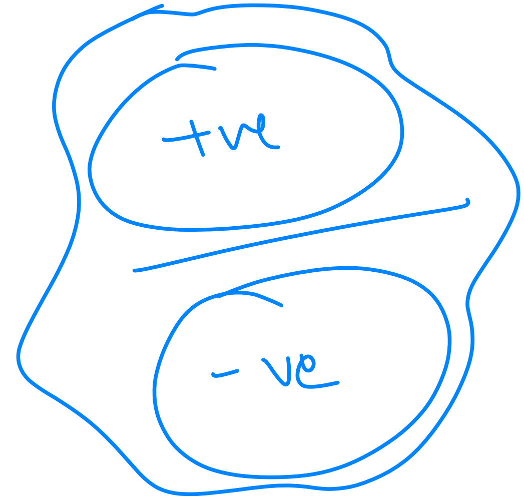
$P(y=3) = 0.3$



$P_1 = 30\%$ $\quad P_2 = 40\%$ $\quad P_3 = 30\%$

$$\mathcal{D} = \begin{cases} \to y = 0 \\ \to y = 1 \end{cases}$$





$Y \to 0$, $Y \to 1$ $\Rightarrow$ Bernoulli

$Y \to 0$, $Y \to 1$, $Y \to 2$, ... $\Rightarrow$ Multinomial

$Y \Rightarrow 1, 2, 3, 4, 5 \ldots k$

$p_1 \quad p_2 \quad p_3 \quad \ldots \quad p_k$

$$\sum_{j=1}^{1} p_k = 1$$

Generate a sample let $(y=1)$

(B)

$$\underbrace{\mu_d^i, \Sigma_{d\times d}^i \quad \forall_i \quad i = 1 \to n}$$

↳ params of each Gaussian

$\therefore$ $Y = 1$

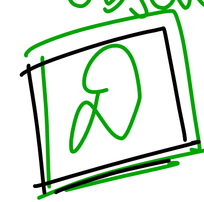
Sample from $N_d(\mu_d^1, \Sigma_{d\times d}^1)$

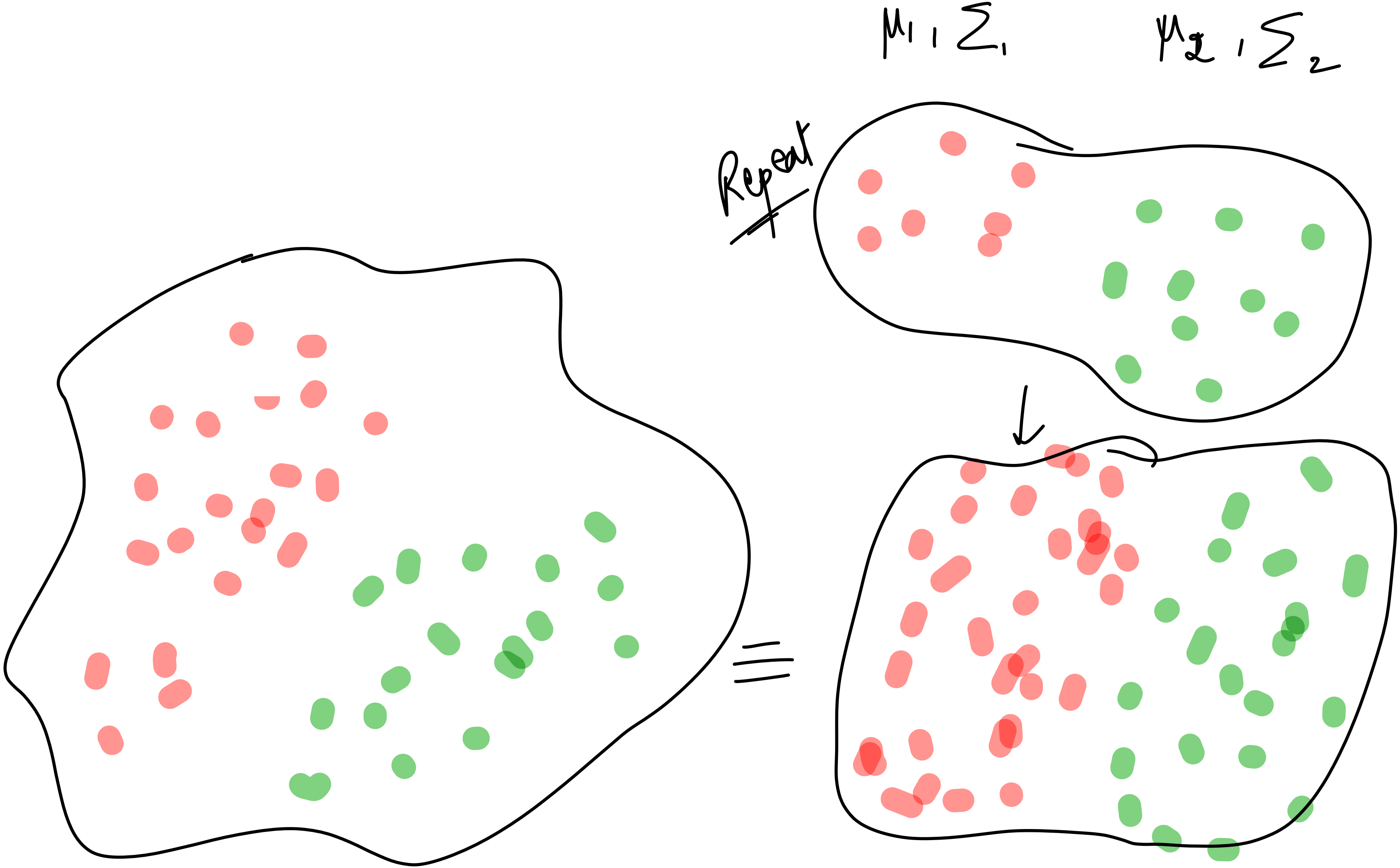
↳ one datapoint.

Repeat (a) & (b) many time

↓

you obtain dataset which is very "similar" to observed




μ_1, Σ_1
μ_2, Σ_2
Repeat
$\equiv$

GMM $\rightarrow$ strong assumption

K - underlying Gaussian

$\llcorner$ Seldom used

Generative model

Find $P(y=j) \; \forall j \; 1{:}k$ $\rightarrow$ param $[\theta]$ (let)

and μ_d, $\Sigma_{d \times d}$

Using $\mathcal{D}$: $\{x_i\}_{i=1}^{n}$ $x_i \in \mathbb{R}^d$

$$\underset{\text{params}}{\Theta} = \left\{ \underbrace{P(y=j)\ \forall j}_{A},\ \underbrace{\mu_d,\ \Sigma_{d\times d}}_{B} \right]$$

Maximum Likelihood Estimation → optimisation

Find θ's s.t. the probability of generating $\mathcal{D}$ is "maximal"

2 steps generative process.

$$\max_{\theta} \; P(\theta)$$

(Generative process)

Underlying GMM

$$\max_{\theta} \; P(x_1, x_2, x_3 \dots)$$

$$D = \{x_i\}_{i=1}^{n}$$

$$\max_{\theta} \; P(x_1 \cap x_2 \cap x_3 \dots \cap x_n)$$

Each x_i is independent of another

$$\max \; P(x_1) \cdot P(x_2) \cdot P(x_3) \cdot \dots$$

$$\max_{\theta} \prod_{i=1}^{n} P(x_i) \rightarrow \text{prob of observing}$$

① j → P_j's

② Generate more data for $N(\mu_d, \Sigma_{d\times d})$

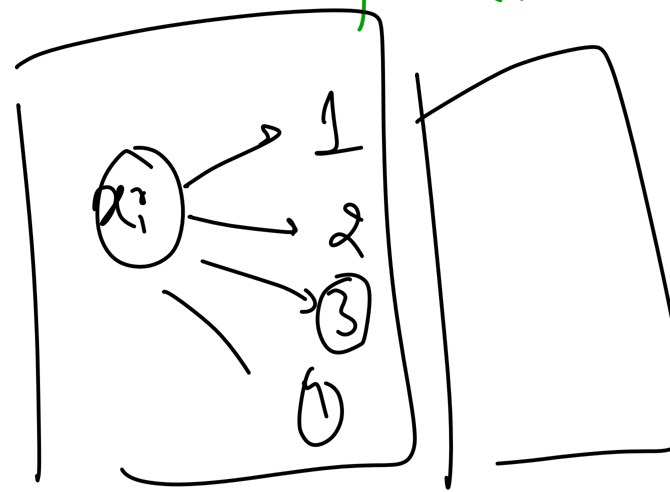
$P(A,B)$

$P(A \cap B)$

$$P(y_j \cap x_i)$$

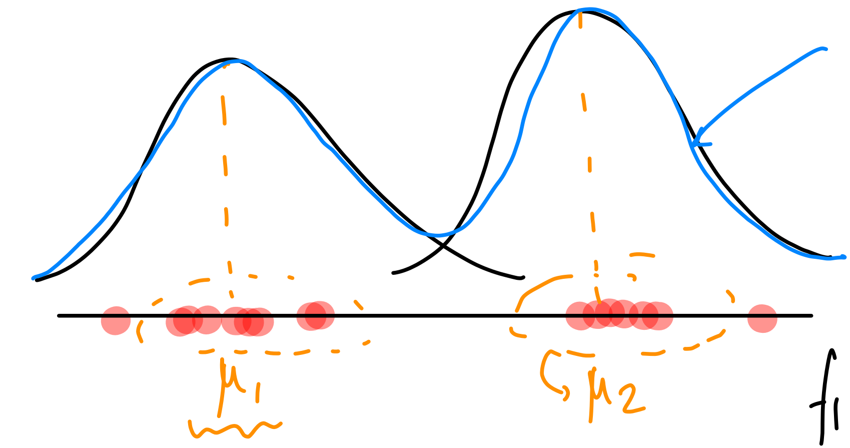
$$\max_{\Theta} \prod_{i=1}^{n} \sum_{j=1}^{k} P(y_j, x_i)$$

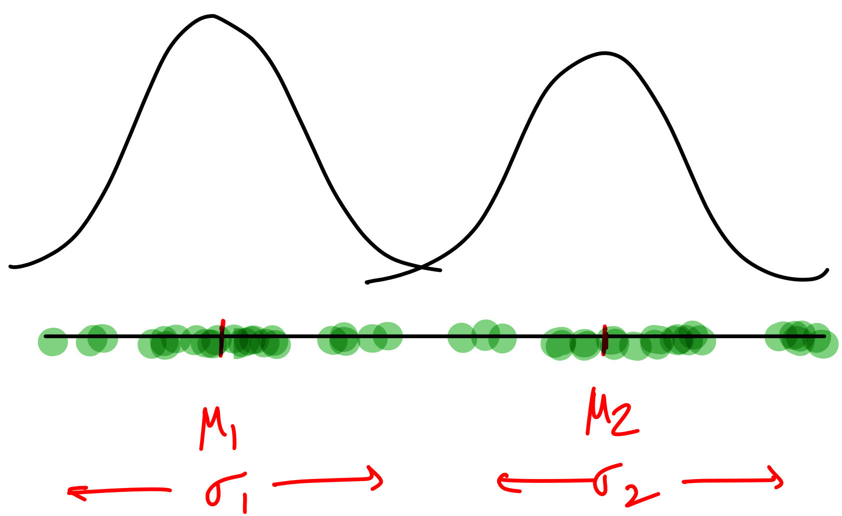
$$\max_{\theta} \prod_{i=1}^{n} \sum_{j=1}^{k} \underbrace{P(x_i \mid y_j)}_{\text{Likelihood fun.}} \cdot \underbrace{P(y_j)}_{\text{Classification}}$$



Param

$\Theta = [P(y=j); \mu_d, \Sigma_{d\times d}]$ for generation


μ_1
μ_2
f_1


μ_1
μ_2
σ_1
σ_2

$$\max_{\theta} \prod_{i=1}^{n} \sum_{j=1}^{k} \underbrace{P(x_i \mid y_i)}_{\text{PDF}} \cdot \underbrace{P(y_i)}_{\text{One of the prior}}$$

PDF

$$(2\pi)^{-k/2}\det(\boldsymbol{\Sigma})^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

One of the prior

μ_d, $\Sigma_{d \times d}$

⇒ ~~Complex~~

⇒ GD - Complex

⇒ Multiple local minima.

"Hack"

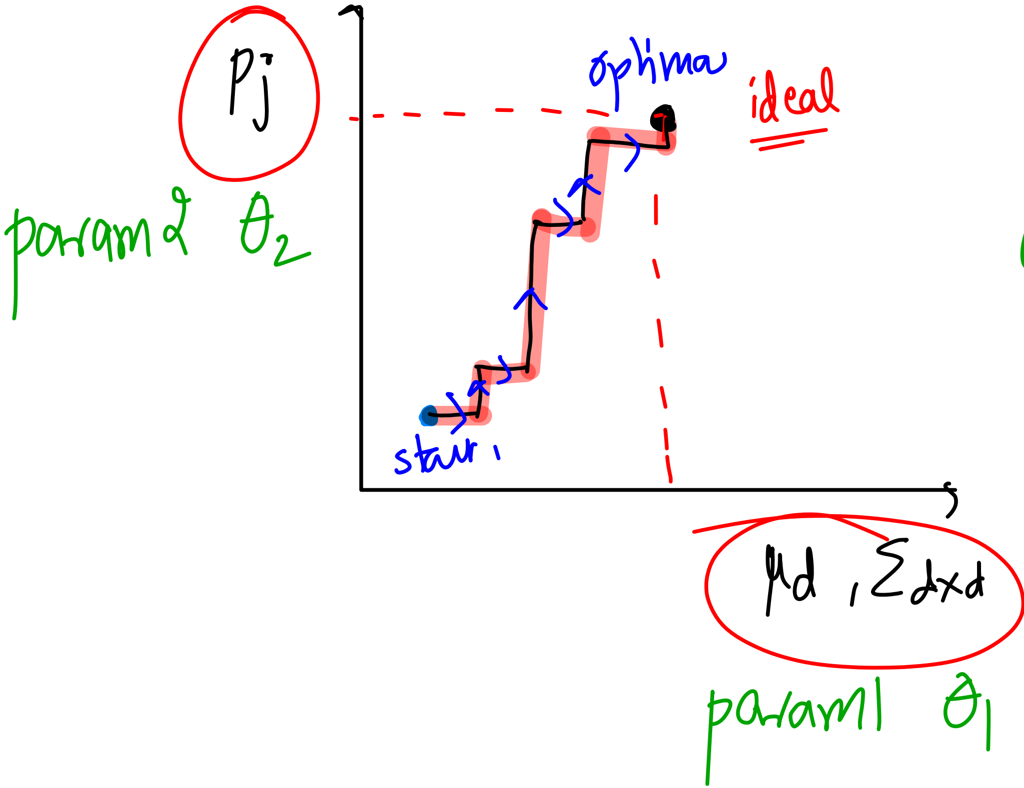
"Expectation - Maximisation"

Core idea

$$\Theta : [\underbrace{p_j}_{A} ; \underbrace{\mu_d, \Sigma_{d \times d}}_{B}]$$

	A	B
E	update	fix
M	fix	update

Computationally cheaper than GD



AKA

Coordinate Ascent