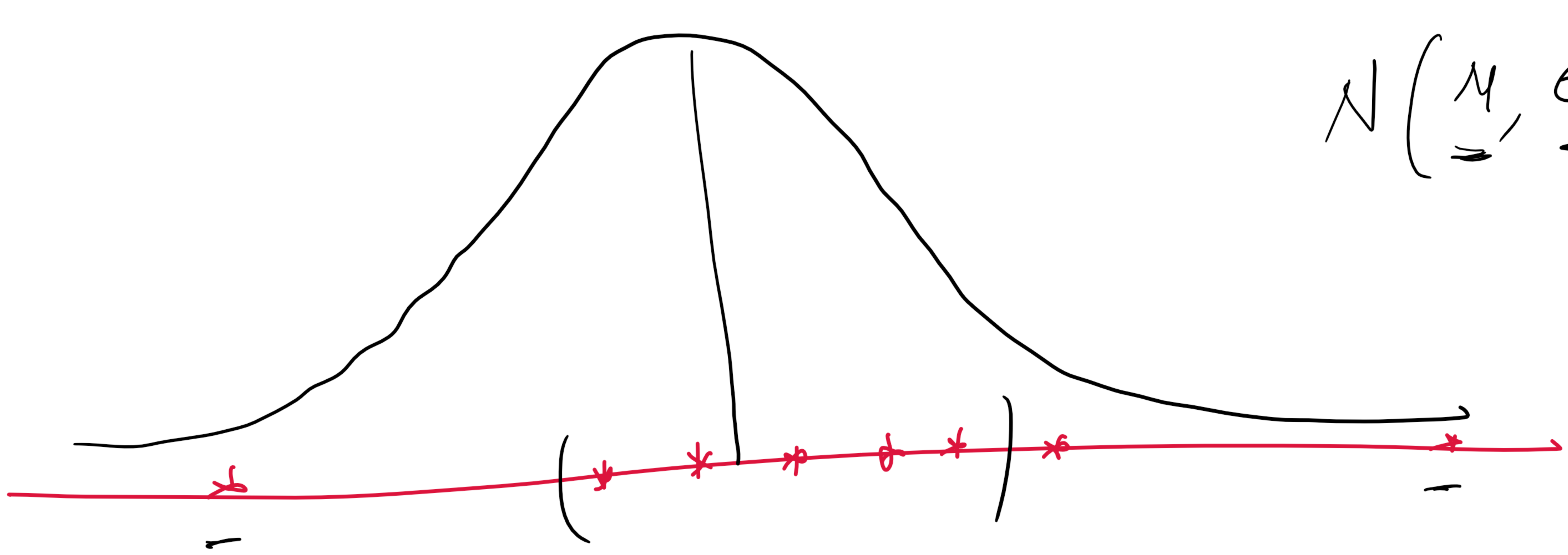
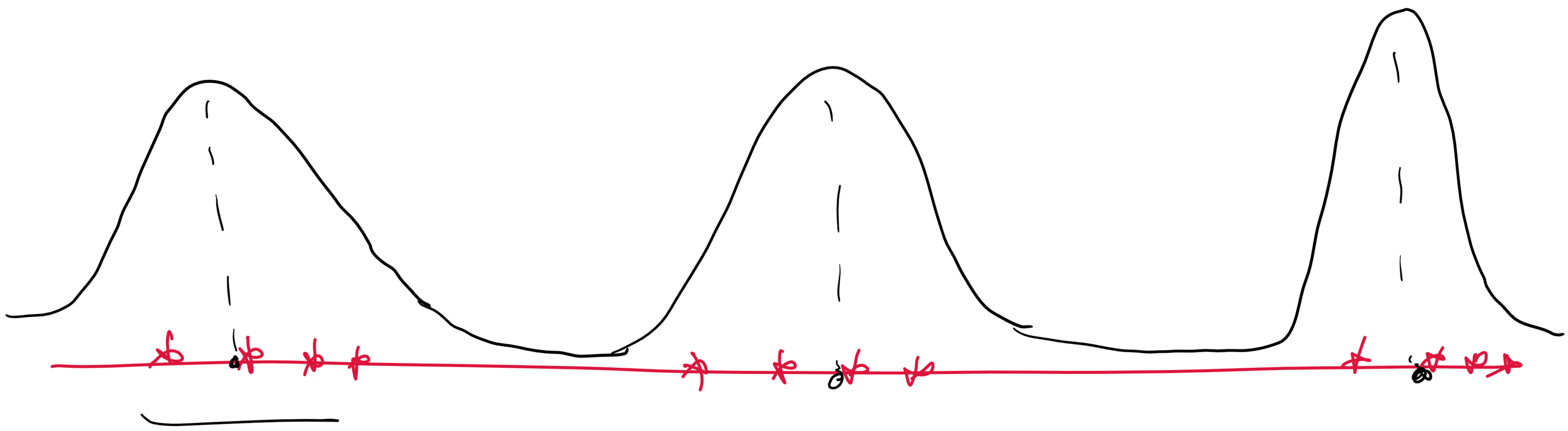
unsupervised learning

Goal: -> Latent variable modeling

$N(\mu, \sigma^2)$



Max Likelihood



MIXTURE OF GAUSSIANS

-> Generate a mixture component among $\{1, \ldots, k\}$ $\quad z_i \in \{1, \ldots, k\}$

$$P(z_i = l) = \pi_l \qquad \boxed{\sum_{i=1}^{k} \pi_i = 1} \qquad 0 \le \pi_i \le 1 \ \forall i$$

Multinomial

$\rightarrow$ Generate $x_i \sim N(\mu_{z_i}, \sigma^2_{z_i})$

$\boxed{\{x_1, \ldots, x_n\} \leftarrow \text{OBSERVED}}$

$z_1, \ldots, z_n$ $\leftarrow$ UNOBSERVED / LATENT

$\in \{1, \ldots, k\}$ $\in \{1, \ldots, k\}$

Parameters to estimate?

$$\left.\begin{array}{l} \mu_1, \ldots, \mu_k \\ \sigma_1^2, \ldots, \sigma_k^2 \\ \pi_1, \pi_2, \ldots, \pi_k \end{array}\right\} \underline{\underline{3k-1}}$$

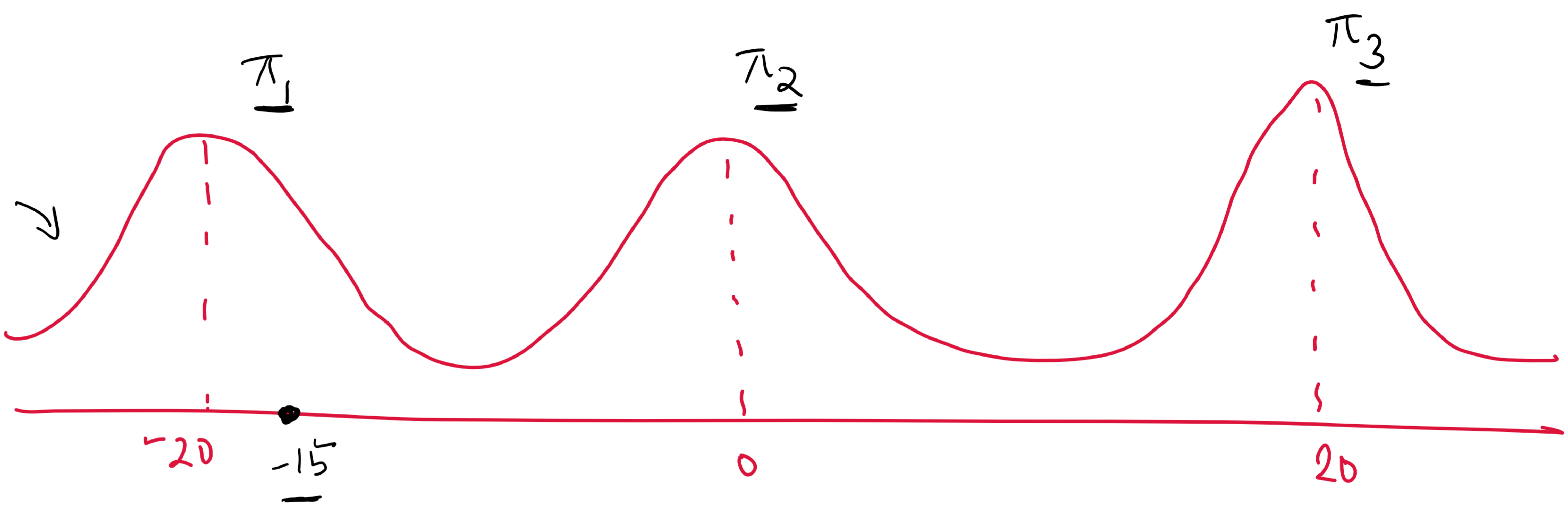
Maximum Likelihood

Mix of Gaussian density. (pdf)

$$L\left(x_1, \ldots, x_n ; \begin{array}{l} \mu_1, \ldots, \mu_k \\ \sigma_1^2, \ldots, \sigma_k^2 \\ \pi_1, \ldots, \pi_k \end{array}\right) = \prod_{i=1}^{n} f_{mix}\left(x_i ; \begin{array}{l} \mu_1, \ldots, \mu_k \\ \sigma_1^2, \ldots, \sigma_k^2 \\ \pi_1, \ldots, \pi_k \end{array}\right)$$

standard normal density

$$= \prod_{i=1}^{n}\left(\sum_{k=1}^{K} \pi_k f(x_i ; \mu_k, \sigma_k^2)\right)$$



$$\pi_1 \, e^{-\frac{(-15-(-20))^2}{2\sigma_1^2}} + \pi_2 \, e^{-\frac{(-15-0)}{2\sigma_2^2}} + \pi_3 \, e^{-\frac{(-15-20)^2}{2\sigma_3^2}}$$

$$L(\theta) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}}$$

$\{x_1, \dots, x_n\}$

$\mu_1, \dots \mu_k$

$\sigma_1^2, \dots \sigma_K^2$

$\pi_1, \dots \pi_K$

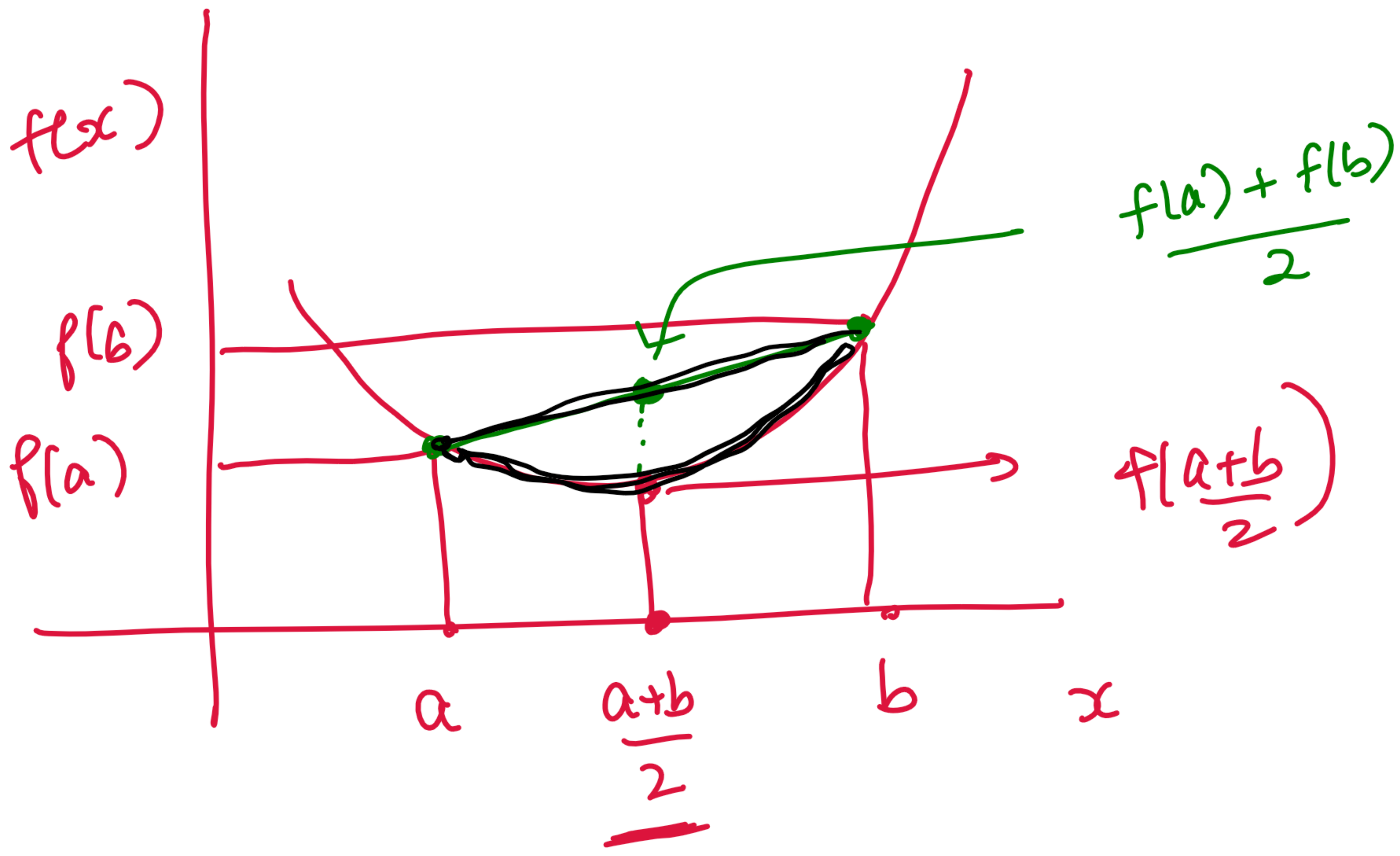
$$\log L(\theta) = \sum_{i=1}^{n} \boxed{\log\left(\sum_{k=1}^{K} \pi_k \, e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}} \frac{1}{\sqrt{2\pi\sigma_k^2}}\right)}$$

can attempt

$$\frac{\partial \log L}{\partial \mu_p} = \sum_{i=1}^{n} \frac{\pi_p \, e^{-\frac{(x_i-\mu_p)^2}{2\sigma_p^2}}}{\sum_{k=1}^{K} \pi_k \, e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}} \frac{1}{\sqrt{2\pi\sigma_k^2}}} \left(-\frac{(x_i-\mu_p)}{2\sigma_p^2}\right) = 0$$

- Not possible to solve analytically
- Need some alternate way to solve this!

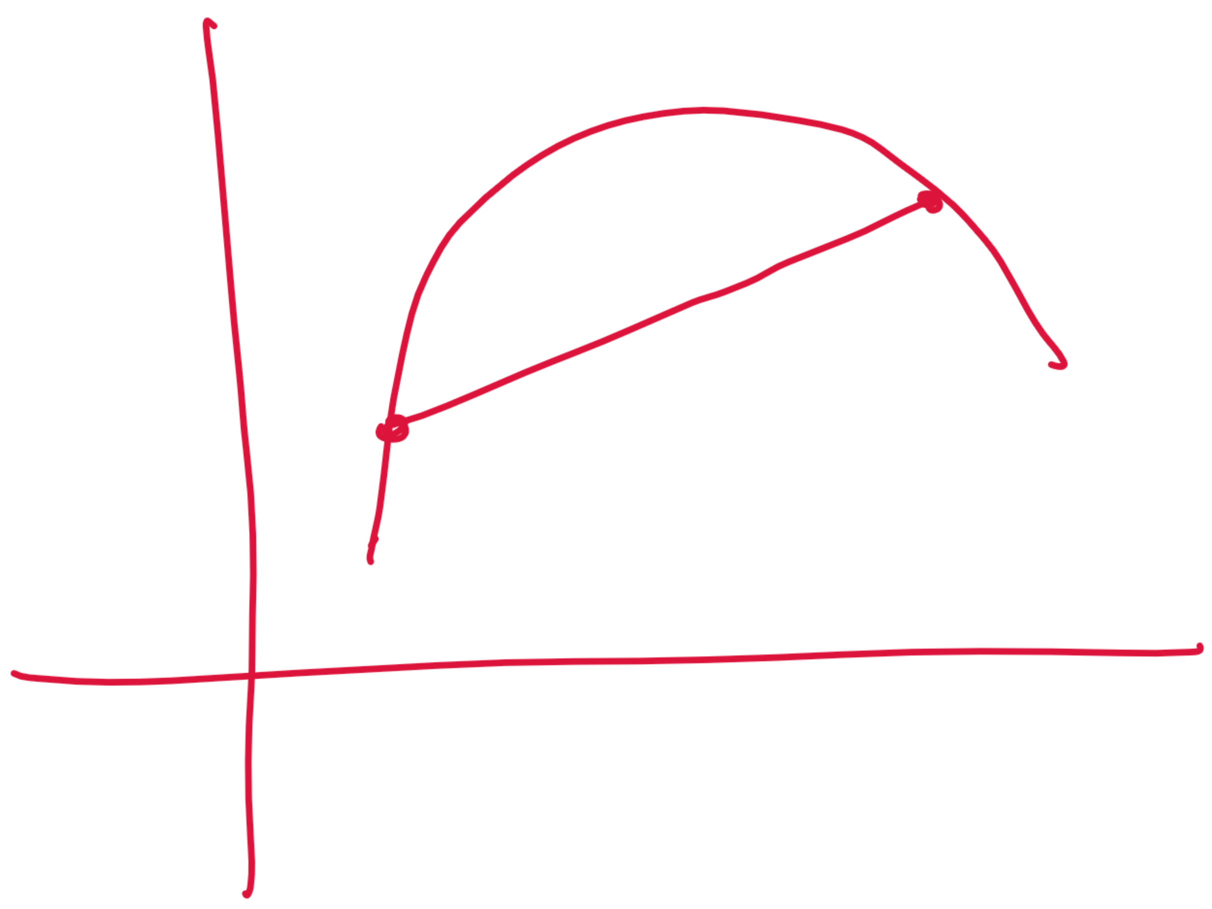
DETOUR : Convex functions



$\forall\, a,b$

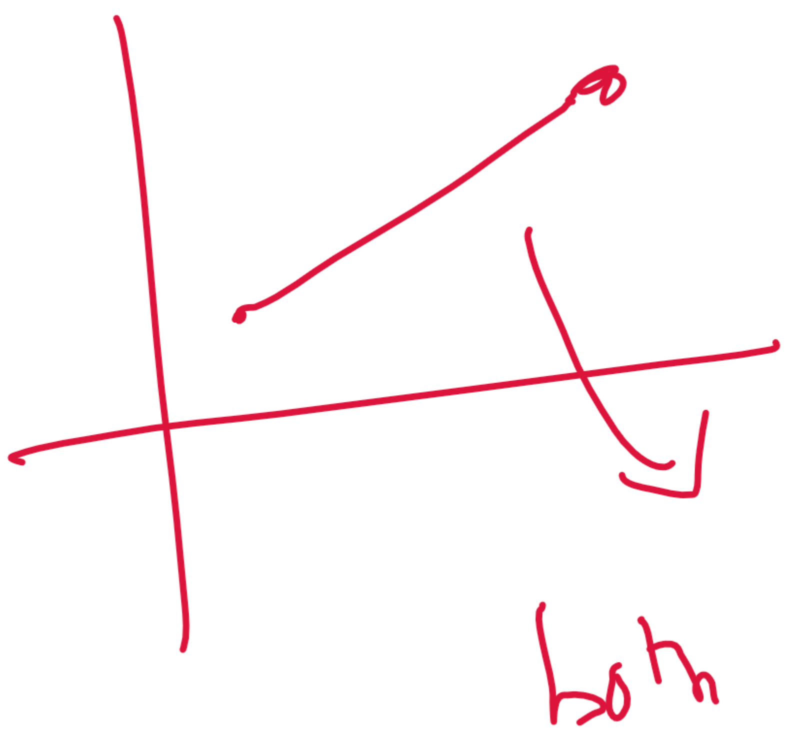
$$f\left(\frac{a+b}{2}\right) \le \frac{f(a)+f(b)}{2}$$

✓ $\Rightarrow$ Convex function.

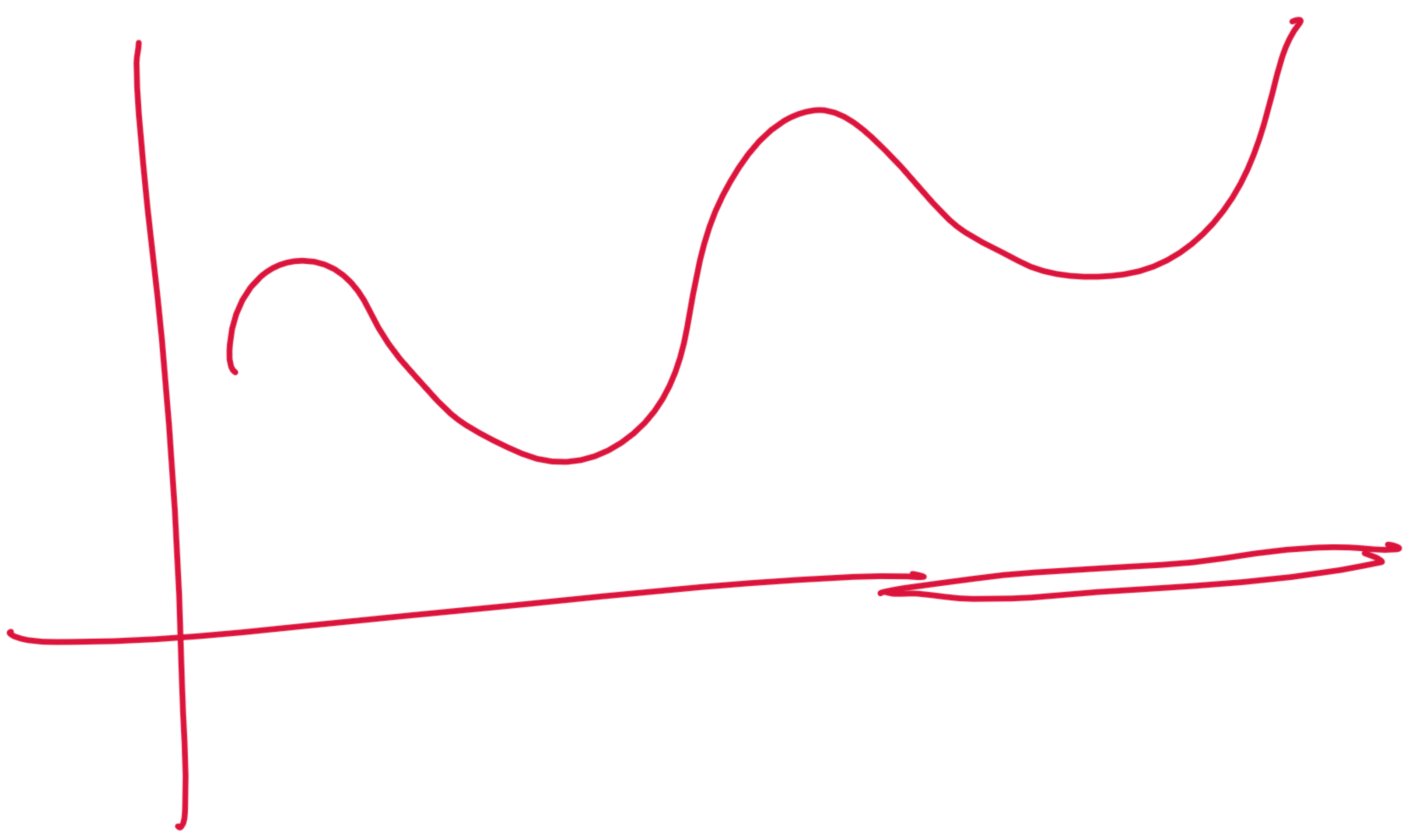


$$f\left(\frac{a+b}{2}\right) \ge \frac{f(a)+f(b)}{2}$$

✓ $\Rightarrow$ concave function



both concave & convex.



$\Rightarrow$ Neither convex nor concave.

$$f\left(\tfrac{1}{2}a + \tfrac{1}{2}b\right) \le \tfrac{1}{2} f(a) + \tfrac{1}{2} f(b)$$

$$\rightarrow \quad f(\lambda a + (1-\lambda) b) \le \lambda f(a) + (1-\lambda) f(b) \qquad \forall \lambda \in [0,1]$$

Concave

$$f(\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_k a_k) \ge \lambda_1 f(a_1) + \dots + \lambda_k f(a_k)$$

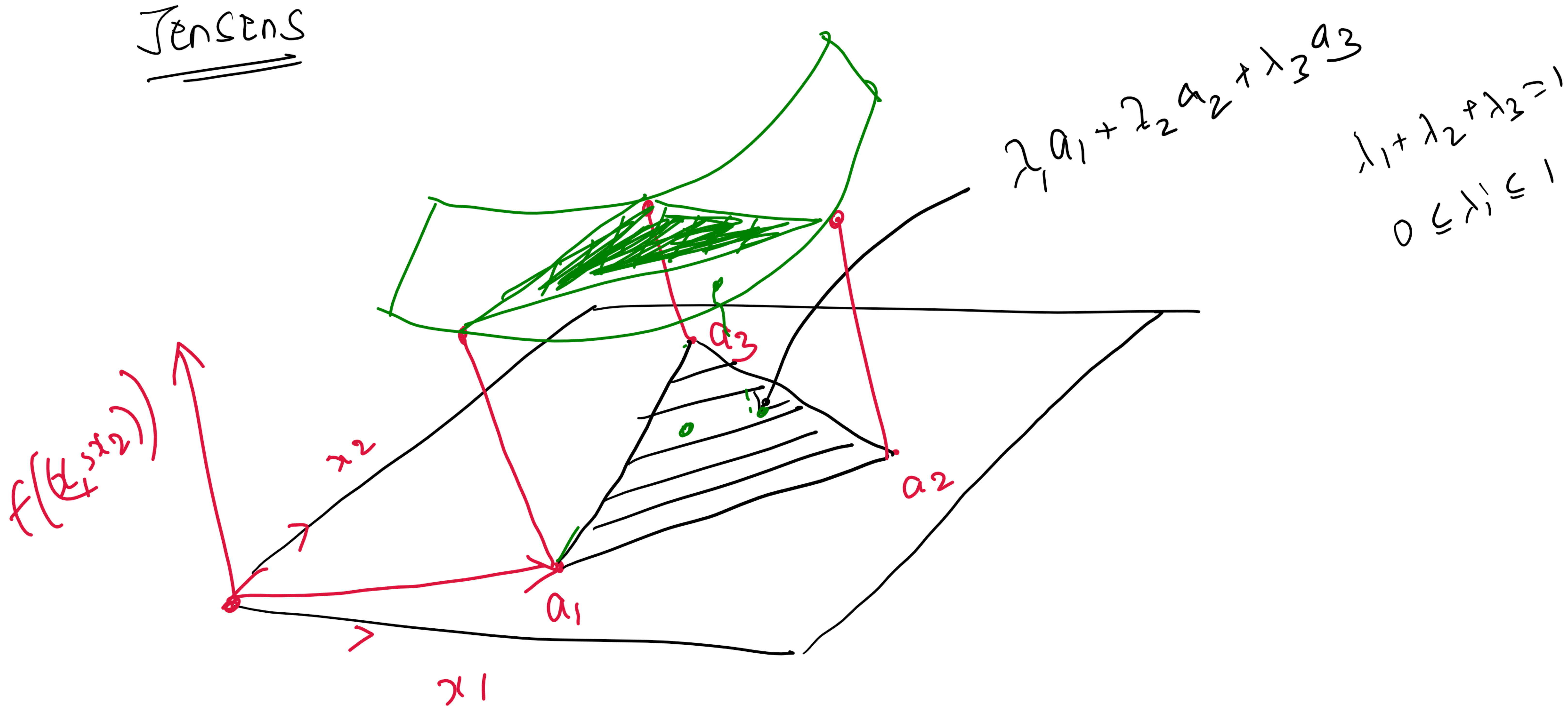
$$\lambda_i \ge 0 \qquad \sum_{i=1}^{k} \lambda_i = 1$$

$$\boxed{f\left(\sum_{k=1}^{k} \lambda_k a_k\right) \ge \sum_{k=1}^{k} \lambda_k f(a_k)}$$

$\hookrightarrow$ JENSEN'S INEQUALITY.

Log is a concave function!

Jensen's



Jensen's for concave functions

$$f\left(\sum_{k=1}^{K} \lambda_k a_k\right) \boxed{\geq} \sum_{k=1}^{K} \lambda_k f(a_k)$$

If f is linear equality holds.

log is concave.

$\lambda_i \geq 0 \qquad \sum_{k=1}^{K} \lambda_k = 1$

$$\log L(x_1, \dots x_n; \theta) = \sum_{i=1}^{n} \log\left(\underline{\sum_{k=1}^{K} \underbrace{P(x_i, z_i = k; \theta)}}\right) \quad \text{— (1)}$$

want to use

$$\log\left(\sum_{k=1}^{K} \lambda_k a_k\right) \geq \sum_{k} \lambda_k \log(a_k)$$

$$(1) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \lambda_k^i \left[\frac{P(x_i, z_i = k; \theta)}{\lambda_k^i}\right]\right)$$

$\sum_{k=1}^{K} \lambda_k^i = 1 \quad \forall i$

$\lambda_k^i \geq 0 \quad \forall i, k$

$$\log L(\theta) \geq \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\lambda_k^i \log(P(x_i, z_i = k; \theta)) - \lambda_k^i \log \lambda_k^i\right] \quad \text{[by Jensens']}$$

$$\log L(\theta) \geq \underbrace{\sum_{i=1}^{n} \sum_{k=1}^{K} \left[\lambda_k^i \log\left(P(x_i, z_i = k; \theta) \right) - \lambda_k^i \log \lambda_k^i \right]}_{\text{Modified log likelihood function}}$$

Fix $\lambda_k^i \quad \forall i, K$ st $\sum_{k=1}^{K} \lambda_k^i = 1 \;,\; \lambda_k^i \geq 0$

$$M\log L(\theta) \; \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\lambda_k^i \log\left(P(z_i = k; \theta) \cdot P(x_i / z_i = k, \theta) \right) - \lambda_k^i \log \lambda_k^i \right]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\lambda_k^i \log\left(\underline{P(z_i = k; \theta}\right) + \lambda_k^i \log \left(\underline{P(x_i / z_i = k; \theta}\right) \underbrace{- \lambda_k^i \log \lambda_k^i}_{} \right]$$

$\pi_1, \ldots, \pi_K$

$\mu_1, \ldots, \mu_K$
$\sigma_1^2, \ldots, \sigma_K^2$

Constant

For GMMs

$$M.\log L(\theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\lambda_k^i \log \pi_k + \boxed{\lambda_k^i} \log \left(\frac{1}{\sqrt{2\pi \sigma_k^2}} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \right) \right]$$

$$\frac{\partial M \cdot \log L}{\partial \mu_l} = 0 \qquad \underset{\text{Mod. Max. Lik}}{\hat{\mu}_l} = \left(\frac{\sum_{i=1}^{n} \lambda_l^i x_i}{\sum_{i=1}^{n} \lambda_l^i} \right)$$

$$\frac{\partial \, \text{Mod.ML}}{\partial \sigma_\ell^2} = 0 \qquad \hat{\sigma}_\ell^2 = \frac{\sum_{i=1}^{n} \lambda_\ell^i (x_i - \hat{\mu}_\ell)^2}{\sum_{i=1}^{n} \lambda_\ell^i}$$

$$\max_{\pi_1, \dots, \pi_k} \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \log \pi_k$$

$$\text{st} \quad \sum_{k=1}^{K} \pi_k = 1 \quad ; \quad \pi_k \geq 0$$

Method of Lagrange multipliers

$$\boxed{\hat{\pi}_\ell = \frac{\sum_{i=1}^{n} \lambda_\ell^i}{n}}$$

$P(z_i = \ell / x_i)$

Algorithm - EM Algorithm (1976)

Expectation
Maximization

- Initialize $\underline{\theta^1}$ ← Iteration

for $t = 1, \ldots, T$

- Find $(\lambda_\ell^i)^t$ by maximizing the modified likelihood with $\theta = \theta^t$

- θ^{t+1} = Maximize the mod. likelihood by fixing $\lambda_\ell^i = (\lambda_\ell^i)^t$

end.

tolerance ϵ

Stopping criterion : $\| \theta^{t+1} - \theta^t \|_2 \leq \epsilon$

Goal: How to maximize $\{\lambda_\ell^i\}$ given θ

$$\max_{\lambda_k^i\ \forall i, \forall k} \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\lambda_k^i \log\left(P(x_i, z_i = k; \theta)\right) - \lambda_k^i \log \lambda_k^i\right]$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \lambda_k^i = 1 \quad \forall i$$

$$0 \le \lambda_k^i \le 1 \quad \forall i, k.$$

Fix some i

$$\left\{ \max_{\lambda_k^i\ \forall k} \sum_{k=1}^{K} \lambda_k^i \left[\log P(x_i, z_i = k; \theta) - \lambda_k^i \log \lambda_k^i\right] \right.$$

$$\text{s.t.} \quad \sum_k \lambda_k^i = 1 \qquad 0 \le \lambda_k^i \le 1 \quad \forall k$$

Solving the Lagrangian for each i

$$\lambda_k^i = P(z_i = k \mid x_i; \theta)$$

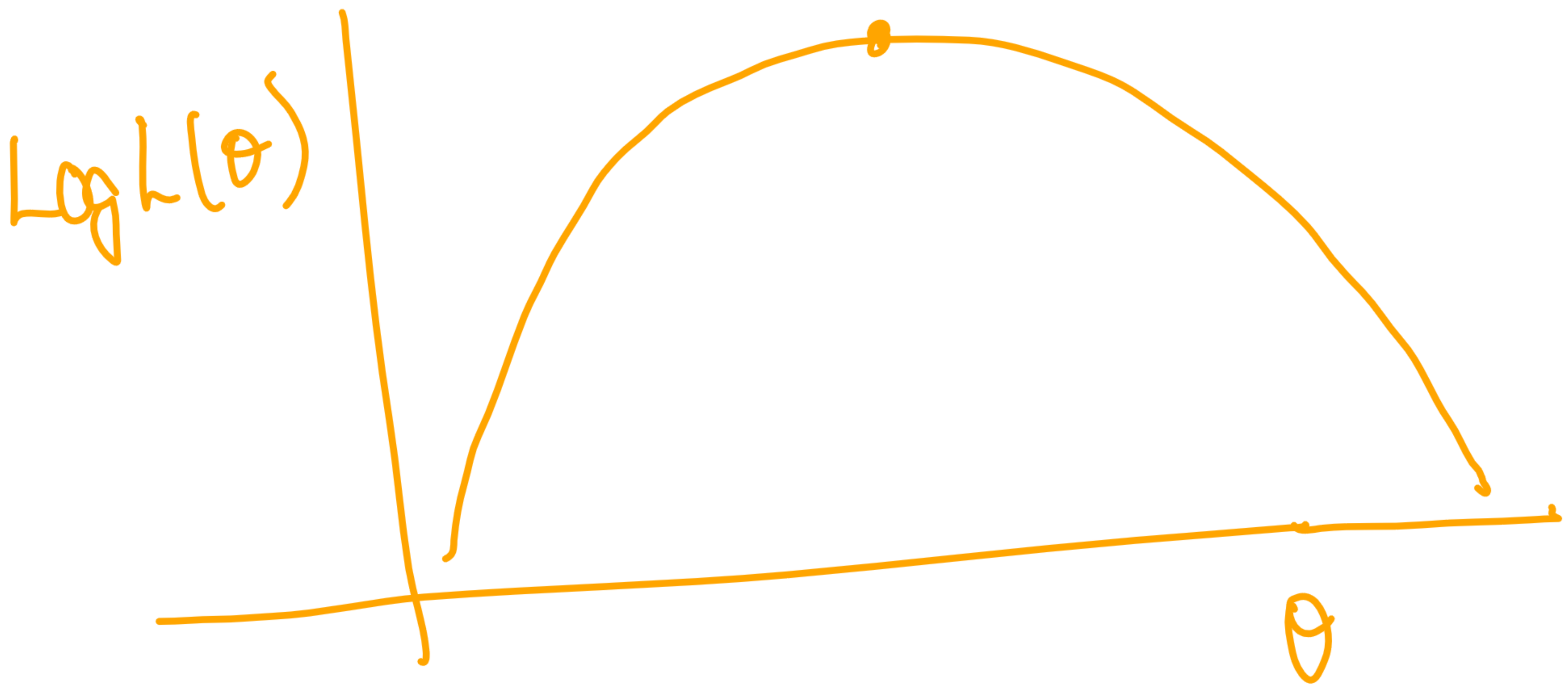
Fixed. $\left[\begin{array}{l} \mu_1, \ldots, \mu_k \\ \sigma_1^2, \ldots, \sigma_k^2 \\ \pi_1, \ldots, \pi_k \end{array}\right.$

π_k

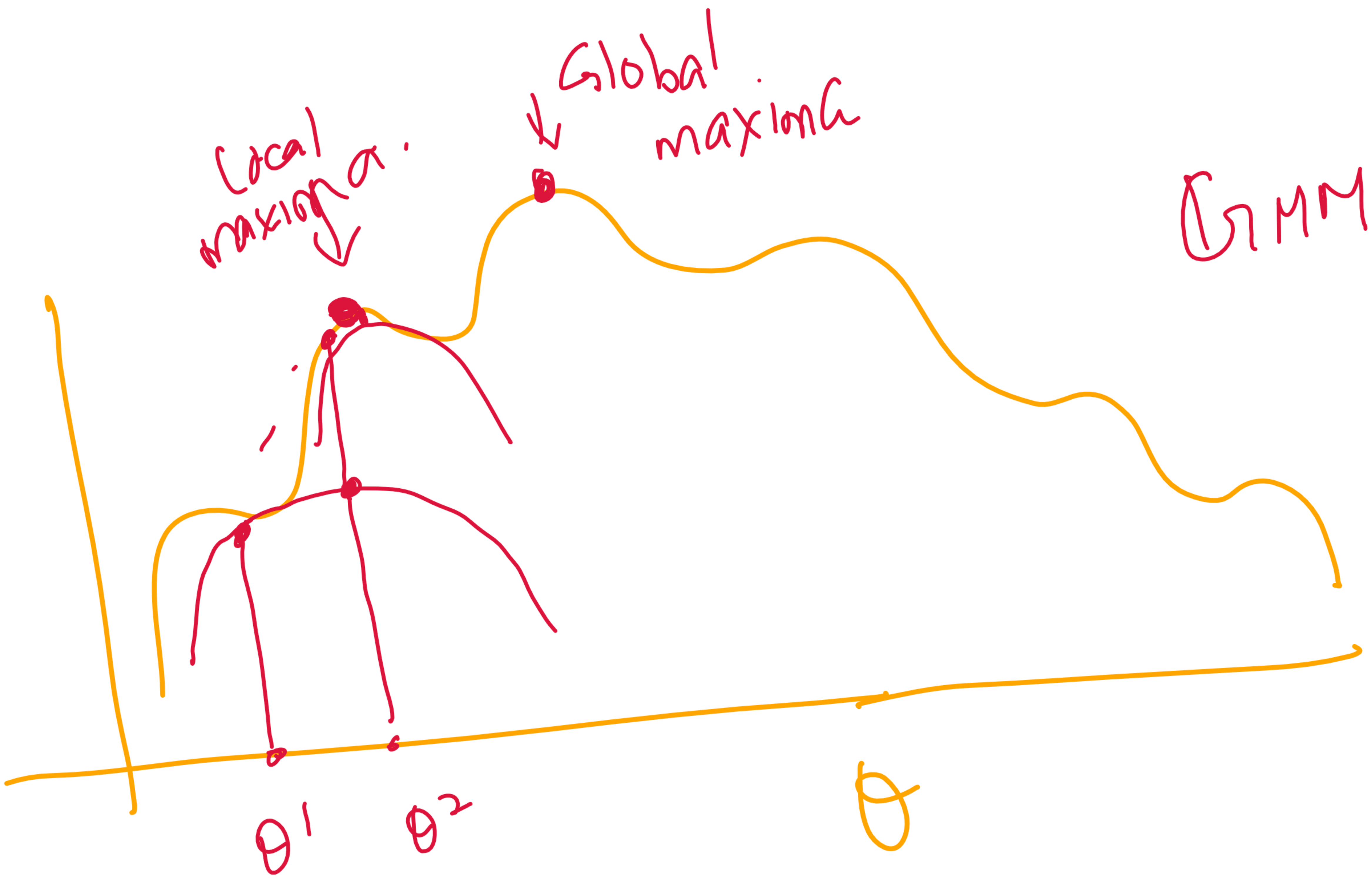
$$\frac{P(z_i = k; \theta) \cdot P(x_i \mid z_i = k; \theta)}{\sum_\ell P(z_i = \ell; \theta) \cdot P(z_i = \ell; \theta)}$$

→ EM algorithm converges

→ It might converge to a "local" maxima.



Standard Gaussian



GMM