CSE6740 09/08/2025 Notes

Name: Junbo Zou
GTID: 904145436

September 8, 2025

1 EM (Expectation–Maximization) Algorithm

Initialize parameters $\{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$. Then:

E-step. Compute the responsibilities

$$\tau_{ik} = p(z_i = k \mid x_i, \theta) = \frac{\pi_k \, \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{k'=1}^K \pi_{k'} \, \mathcal{N}(x_i \mid \mu_{k'}, \Sigma_{k'})}, \quad i = 1, \dots, n, \;\; k = 1, \dots, K.$$

M-step. Update parameters

$$\pi_k = \frac{1}{n} \sum_{i=1}^n \tau_{ik}, \qquad \mu_k = \frac{\sum_{i=1}^n \tau_{ik} \, x_i}{\sum_{i=1}^n \tau_{ik}}, \qquad \Sigma_k = \frac{\sum_{i=1}^n \tau_{ik} \, (x_i - \mu_k)(x_i - \mu_k)^\top}{\sum_{i=1}^n \tau_{ik}}.$$

Log-likelihood. EM maximizes the data log-likelihood

$$\ell(\theta; \mathcal{D}) = \log \prod_{i=1}^n \sum_{z_i=1}^K p(x_i, z_i \mid \theta).$$

Variational interpretation.

E-step: $\ell(\theta; \mathcal{D}) \;\geq\; \ell(\theta, \mathcal{D}; q) = \mathbb{E}_{z_{1:n} \sim q}\left[\log \prod_{i=1}^n p(x_i, z_i \mid \theta)\right] + H(q), \qquad H(q) = -\,\mathbb{E}_q[\log q],$

M-step: $\theta^{(t+1)} = \arg\max_\theta \; \ell(\theta, \mathcal{D}; q).$

2 Convex / Concave Functions and Jensen's Inequality

2.1 Convex Sets

A set $A \subseteq \mathbb{R}^n$ is convex if

$$\forall\, x, y \in A, \;\; 0 \leq \alpha \leq 1 \quad \Rightarrow \quad \alpha x + (1 - \alpha) y \in A.$$

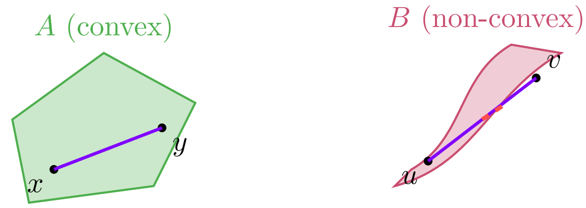

Figure 1: Left: a convex set A (every chord lies inside). Right: a non-convex set B (some chords exit the set).

A set C is a convex cone if

$$\forall\, x_1, x_2 \in C,\ \ \theta_1, \theta_2 \geq 0 \quad \Rightarrow \quad \theta_1 x_1 + \theta_2 x_2 \in C.$$

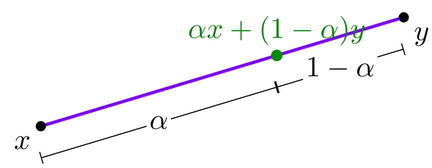

Figure 2: Any convex combination $\alpha x + (1-\alpha)y$ with $\alpha \in [0,1]$ lies on the segment $[x, y]$.

A hyperplane has the form

$$\{\, x \in \mathbb{R}^n \mid a^\top x - x_0 = 0,\ a \neq 0 \,\}.$$

A halfspace is

$$\{\, x \in \mathbb{R}^n \mid a^\top x - x_0 \leq 0,\ a \neq 0 \,\}.$$

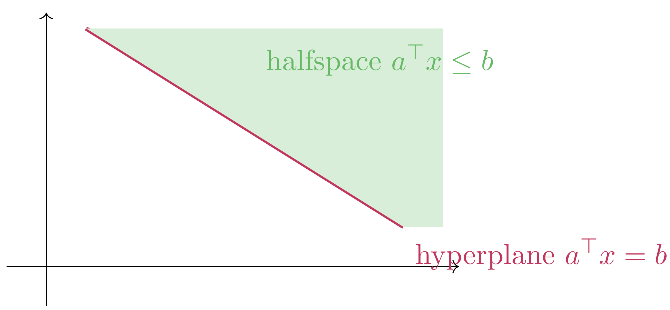

Figure 3: A hyperplane splits $\mathbb{R}^n$ into two convex halfspaces.

A Euclidean ball is

$$B(x_c, r) = \{\, x \in \mathbb{R}^n \mid \|x - x_c\|_2 \leq r \,\}.$$

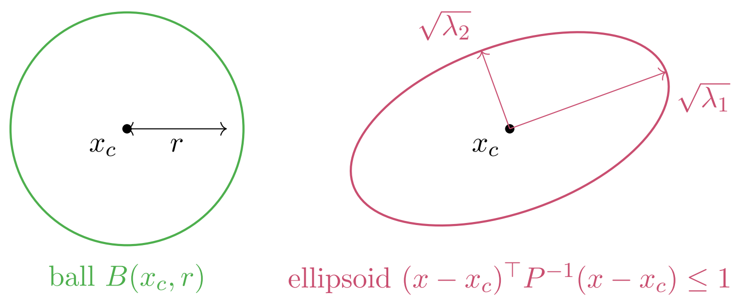

Figure 4: Both balls and ellipsoids are convex sets; ellipsoid axes relate to $P = Q\Lambda Q^\top$.

An ellipsoid is

$$E = \{\, x \in \mathbb{R}^n \mid (x - x_c)^\top P^{-1}(x - x_c) \le 1 \,\}, \quad P \succ 0.$$

If $P = Q\Lambda Q^\top$ with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then the principal directions are columns of Q, and semi-axis lengths are $\sqrt{\lambda_j}$.

A polyhedron is the intersection of finitely many halfspaces/hyperplanes:

$$P = \{\, x \in \mathbb{R}^n \mid a_j^\top x \le b_j,\ j = 1, \ldots, m;\ \ c_k^\top x = d_k,\ k = 1, \ldots, p\}.$$

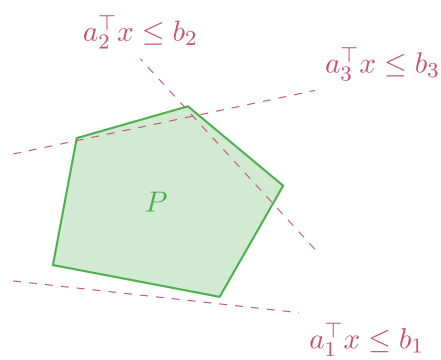

Figure 5: A polyhedron $P = \{x \mid Ax \le b,\ Cx = d\}$ is an intersection of finitely many halfspaces/hyperplanes.

2.2 Convex / Concave Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(\theta x + (1 - \theta)y) \le \theta f(x) + (1 - \theta)f(y), \quad \forall\, x, y,\ \ 0 \le \theta \le 1.$$

It is concave if the inequality is reversed.

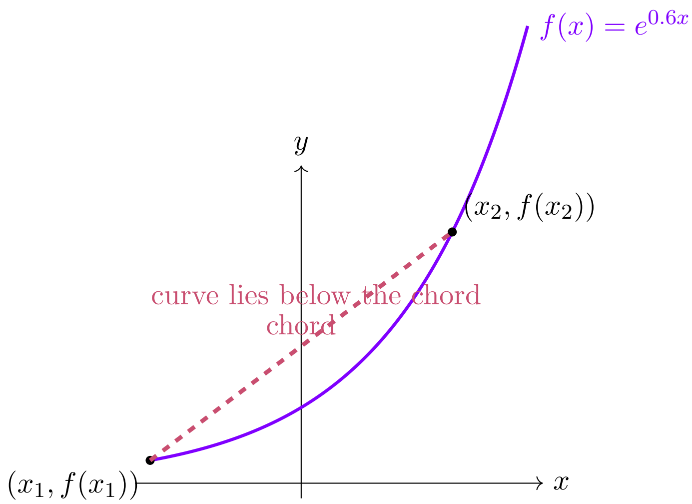

Figure 6: A convex exponential: for any x_1, x_2 and $\theta \in [0,1]$, $f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta)f(x_2)$.

2.3 Conditions for Convexity

First-order condition (differentiable):

$$f(y) \;\ge\; f(x) + \nabla f(x)^\top (y-x), \quad \forall\, x, y \in \mathrm{dom}(f).$$

Second-order condition (twice differentiable):

$$\nabla^2 f(x) \succeq 0, \quad \forall\, x \in \mathrm{dom}(f).$$

Example. Quadratic form $f(x) = \frac{1}{2}x^\top A x$ is convex iff $A \succeq 0$.

2.4 Jensen's Inequality

For concave f:

$$f\Big(\sum_{i=1}^m a_i x_i\Big) \;\ge\; \sum_{i=1}^m a_i f(x_i), \qquad a_i \ge 0,\; \sum_{i=1}^m a_i = 1.$$

In expectation form:

$$f(\mathbb{E}[X]) \;\ge\; \mathbb{E}[f(X)].$$

3 Expectation Step Solution

Let $x \in \mathbb{R}^d$ and a discrete latent $z \in \{1, \dots, K\}$. For a single observation,

$$\begin{aligned}\ell(\theta; x) &= \log \sum_{z=1}^K p(x, z \mid \theta) = \log \sum_{z=1}^K q(z)\, \frac{p(x,z \mid \theta)}{q(z)} \\ &\ge \sum_{z=1}^K q(z)\, \log \frac{p(x,z \mid \theta)}{q(z)} = \mathbb{E}_{z \sim q}\big[\log p(x, z \mid \theta)\big] + H(q),\end{aligned}$$

where $H(q) = -\sum_z q(z)\log q(z)$ and the inequality is Jensen's (concave log).

For a dataset $D = \{x_i\}_{i=1}^n$, we have

$$\begin{aligned}\ell(\theta; D) &\geq \sum_{i=1}^{n} \Big(\mathbb{E}_{z_i\sim q_i}[\log p(x_i, z_i \mid \theta)] + H(q_i)\Big) \\ &= \mathbb{E}_{q(z_{1:n})}\Big[\log \prod_{i=1}^{n} p(x_i, z_i \mid \theta)\Big] + H(q) \;\equiv\; \mathcal{L}(\theta, D; q).\end{aligned}$$

E-step (tightness of the bound). Choose $q_i(z = k) = p(z_i = k \mid x_i, \theta) \equiv \tau_i^k$. Then

$$\frac{p(x, z \mid \theta)}{q(z)} = \frac{p(z, x \mid \theta)}{p(z \mid x, \theta)} = p(x \mid \theta),$$

which is independent of z, so the inequality becomes equality and $\mathcal{L}(\theta, D; q) = \ell(\theta; D)$ at the chosen q.

E-step (computing the expectation). For GMM, $p(x_i, z_i = k \mid \theta) = \pi_k \mathcal{N}(x_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. With $\tau_i^k = p(z_i = k \mid x_i, \theta)$,

$$\begin{aligned}\mathcal{L}(\theta; D, q) &= \sum_{i=1}^{n}\sum_{k=1}^{K} \tau_i^k \log\big(\pi_k \mathcal{N}(x_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\big) + H(q) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} \tau_i^k \Big[\log \pi_k - \tfrac{1}{2}(x_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(x_i - \boldsymbol{\mu}_k) - \tfrac{1}{2}\operatorname{logdet} \boldsymbol{\Sigma}_k - c\Big] + H(q),\end{aligned}$$

where $c = \frac{d}{2}\log(2\pi)$ is constant w.r.t. θ.

M-step (maximize $\mathcal{L}$ w.r.t. θ). Maximizing over $\{\pi_k\}$ with $\sum_k \pi_k = 1$ gives

$$\boxed{\pi_k^{\text{new}} = \frac{1}{n}\sum_{i=1}^{n} \tau_i^k}$$

and maximizing over $\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ yields the weighted MLEs

$$\boxed{\boldsymbol{\mu}_k^{\text{new}} = \frac{\sum_{i=1}^n \tau_i^k\, x_i}{\sum_{i=1}^n \tau_i^k}, \qquad \boldsymbol{\Sigma}_k^{\text{new}} = \frac{\sum_{i=1}^n \tau_i^k\,(x_i - \boldsymbol{\mu}_k^{\text{new}})(x_i - \boldsymbol{\mu}_k^{\text{new}})^\top}{\sum_{i=1}^n \tau_i^k}}.$$

EM updates (for completeness). With current parameters θ,

$$\textbf{E-step:}\quad \tau_i^k = \frac{\pi_k \mathcal{N}(x_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k'=1}^K \pi_{k'} \mathcal{N}(x_i \mid \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})}, \qquad i = 1, \ldots, n;\; k = 1, \ldots, K,$$

$$\textbf{M-step:}\quad \pi_k^{\text{new}} = \frac{1}{n}\sum_i \tau_i^k, \quad \boldsymbol{\mu}_k^{\text{new}} = \frac{\sum_i \tau_i^k\, x_i}{\sum_i \tau_i^k}, \quad \boldsymbol{\Sigma}_k^{\text{new}} = \frac{\sum_i \tau_i^k (x_i - \boldsymbol{\mu}_k^{\text{new}})(x_i - \boldsymbol{\mu}_k^{\text{new}})^\top}{\sum_i \tau_i^k}.$$

Repeat E/M until convergence (the data log-likelihood is non-decreasing).