Approximate Bayesian inference

Ekaterina Lobacheva
Research Fellow at Samsung-HSE Laboratory

DeepBayes

Approximate Bayes: problem set

The problem set is available here:

tiny.cc/AB_problems

Clustering problem

Clustering problem:

- Dataset $X = \{x_i\}_{i=1}^N$
- We want to group these objects into K clusters

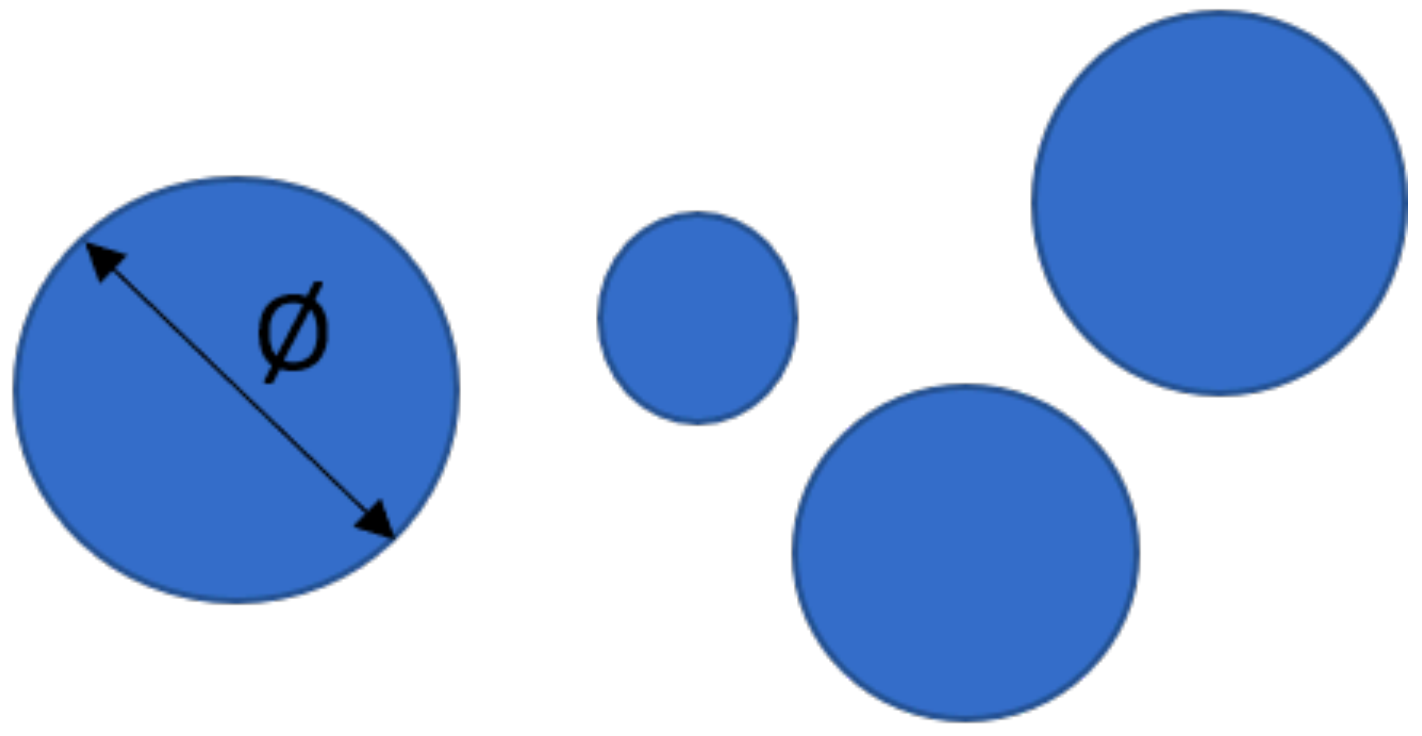
Gaussian mixture model:

- K Gaussian components with probabilities $\pi = (\pi_1, \ldots, \pi_K)$
- Each Gaussian has parameters μ_k, λ_k
- Each object has latent variable which shows affiliation to a cluster:

$$z_i \in \{0, 1\}^K, \quad \sum_{k=1}^K z_{ik} = 1$$

Clustering problem

K-means



- Hard clustering
- Diagonal covariance matrices

GMM

- Soft clustering
- Trainable covariance matrices
- We may use priors for π, μ, λ

Gaussian mixture model

Basic probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

We want to train the model:

- Find posterior distribution for latent variables
- Find optimal values for parameters

Inference methods

Only parameters → Maximum likelihood

Inference methods

Only parameters → Maximum likelihood

Only latent variables →

Conjugacy	Full Bayesian inference
Conditional conjugacy	Mean field variational inference
No conjugacy	Parametric variational inference

Inference methods

Only parameters → Maximum likelihood

Only latent variables →

Conjugacy	Full Bayesian inference
Conditional conjugacy	Mean field variational inference
No conjugacy	Parametric variational inference

Parameters and latent variables → + EM-algorithm: a suitable method for latent variables on E-step and maximisation on M-step

Problem 1: basic GMM

Probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) = \prod_{i=1}^{N}\prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

How to train?

Problem 1: basic GMM

Probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) = \prod_{i=1}^{N}\prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

How to train? EM-algorithm

- E-step: find posterior distribution for latent variables Z
- M-step: find optimal values for parameters π, μ, λ

Problem 1: basic GMM

Probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) = \prod_{i=1}^{N}\prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are conjugate.
- E-step: derive $p(Z|X, \pi, \mu, \lambda)$. Values of π, μ, λ are fixed on this step.
- M-step: compute optimal values of π, μ, λ by maximizing $\mathbb{E}_{p(Z|X,\pi,\mu,\lambda)} \log p(X, Z|\pi, \mu, \lambda)$. Posterior distribution on Z is fixed on this step.

Problem 1: basic GMM

Probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) = \prod_{i=1}^{N}\prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are conjugate.
- E-step: derive $p(Z|X, \pi, \mu, \lambda)$. Values of π, μ, λ are fixed on this step.
- M-step: compute optimal values of π, μ, λ by maximizing $\mathbb{E}_{p(Z|X,\pi,\mu,\lambda)} \log p(X, Z|\pi, \mu, \lambda)$. Posterior distribution on Z is fixed on this step.

Problem 1: Gaussian mixture model

Probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 1: Gaussian mixture model

Probabilistic model: $$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Prior: $$p(Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{ik}} = \prod_{i=1}^{N} \prod_{k=1}^{K} C^{z_{ik}}$$

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 1: Gaussian mixture model

Probabilistic model: $$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Prior: $$p(Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{ik}} = \prod_{i=1}^{N} \prod_{k=1}^{K} C^{z_{ik}}$$

Posterior: $$p(Z|X) \propto p(X, Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}} = \prod_{i=1}^{N} \prod_{k=1}^{K} C^{z_{ik}}$$

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 1: Gaussian mixture model

Probabilistic model: $$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Prior: $$p(Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{ik}} = \boxed{\prod_{i=1}^{N} \prod_{k=1}^{K} C^{z_{ik}}}$$

conjugate

Posterior: $$p(Z|X) \propto p(X, Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}} = \boxed{\prod_{i=1}^{N} \prod_{k=1}^{K} C^{z_{ik}}}$$

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 1: basic GMM

Probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) = \prod_{i=1}^{N}\prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are conjugate.
- E-step: derive $p(Z|X, \pi, \mu, \lambda)$. Values of π, μ, λ are fixed on this step.
- M-step: compute optimal values of π, μ, λ by maximizing $\mathbb{E}_{p(Z|X,\pi,\mu,\lambda)} \log p(X, Z|\pi, \mu, \lambda)$. Posterior distribution on Z is fixed on this step.

Problem 1: Bayesian framework

Likelihood and prior are conjugate $\longrightarrow$ posterior is Categorical

$$p(Z|X) \propto p(X,Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Problem 1: Bayesian framework

Likelihood and prior are conjugate $\longrightarrow$ posterior is Categorical

$$p(Z|X) \propto p(X,Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \boxed{\left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}}$$

$$p(z_{ik} = 1|X) \propto \pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})$$

Problem 1: Bayesian framework

Likelihood and prior are conjugate $\longrightarrow$ posterior is Categorical

$$p(Z|X) \propto p(X, Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \boxed{\left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}}$$

Normalization

$$\boxed{\sum_{k=1}^{K} p(z_{ik} = 1|X) = 1}$$

$$p(z_{ik} = 1|X) \propto \pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})$$

Problem 1: Bayesian framework

Likelihood and prior are conjugate $\longrightarrow$ posterior is Categorical

$$p(Z|X) \propto p(X,Z) = \prod_{i=1}^{N} \prod_{k=1}^{K} \boxed{\left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}}$$

Normalization

$$\boxed{\sum_{k=1}^{K} p(z_{ik}=1|X) = 1}$$

$$p(z_{ik}=1|X) = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \lambda_j^{-1})}$$

Problem 1: Bayesian framework

Likelihood and prior are conjugate $\longrightarrow$ posterior is Categorical

$$p(Z|X) \propto p(X,Z) = \prod_{i=1}^{N}\prod_{k=1}^{K} \boxed{\left[\pi_k \mathcal{N}(x_i|\mu_k,\lambda_k^{-1})\right]^{z_{ik}}}$$

Normalization

$$\boxed{\sum_{k=1}^{K} p(z_{ik}=1|X) = 1}$$

$$p(z_{ik}=1|X) = \frac{\pi_k \mathcal{N}(x_i|\mu_k,\lambda_k^{-1})}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j,\lambda_j^{-1})}$$

Result: $$p(Z|X) = \prod_{i=1}^{N}\prod_{k=1}^{K}\left[\frac{\pi_k \mathcal{N}(x_i|\mu_k,\lambda_k^{-1})}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j,\lambda_k^{-1})}\right]^{z_{ik}}$$

Problem 1: basic GMM

Probabilistic model:

$$p(X, Z|\pi, \mu, \lambda) = \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) = \prod_{i=1}^{N}\prod_{k=1}^{K}\left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are conjugate.
- E-step: derive $p(Z|X, \pi, \mu, \lambda)$. Values of π, μ, λ are fixed on this step.
- M-step: compute optimal values of π, μ, λ by maximizing $\mathbb{E}_{p(Z|X,\pi,\mu,\lambda)} \log p(X, Z|\pi, \mu, \lambda)$. Posterior distribution on Z is fixed on this step.

Problem 1: basic GMM

Rewrite $\mathbb{E}_{p(Z|X)} \log p(X, Z)$ as a function of π, μ, λ:

$$\mathbb{E}_{p(Z|X)} \log p(X, Z) = \mathbb{E}_{p(Z|X)} \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left[\log \pi_k + \log \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right] =$$

$$= \mathbb{E}_{p(Z|X)} \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left[\log \pi_k + \frac{1}{2} \log \lambda_k - \frac{1}{2}(x_i - \mu_k)^2 \lambda_k\right] + C =$$

$$= \left\{\mathbb{E}_{p(Z|X)} z_{ik} = p(z_{ik} = 1|X) = \gamma_{ik}\right\} =$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} \gamma_{ik} \left[\log \pi_k + \frac{1}{2} \log \lambda_k - \frac{1}{2}(x_i - \mu_k)^2 \lambda_k\right] + C$$

Problem 1: basic GMM

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\gamma_{ik}\left[\log\pi_k + \frac{1}{2}\log\lambda_k - \frac{1}{2}(x_i-\mu_k)^2\lambda_k\right] \to \min_{\pi,\mu,\lambda}$$

Optimization w.r.t. π which is restricted to a simplex:

- change parameterization to $\eta_k = \log\pi_k, \quad \eta_k \in \mathbb{R}$
- Lagrangian:

$$\mathcal{L}(\eta,\psi) = \sum_{i=1}^{N}\sum_{k=1}^{K}\gamma_{ik}\eta_k - \psi(\sum_{k=1}^{K}\exp\eta_k - 1)$$

Problem 1: basic GMM

$$\mathcal{L}(\eta, \psi) = \sum_{i=1}^{N} \sum_{k=1}^{K} \gamma_{ik} \eta_k - \psi(\sum_{k=1}^{K} \exp \eta_k - 1)$$

Differentiation:

$$0 = \frac{\partial \mathcal{L}(\eta, \psi)}{\partial \eta_k} = \sum_{i=1}^{N} \gamma_{ik} - \psi \exp \eta_k \Rightarrow \pi_k = \exp \eta_k = \frac{\sum_{i=1}^{N} \gamma_{ik}}{\psi}$$

$$0 = \frac{\partial \mathcal{L}(\eta, \psi)}{\partial \psi} = -\sum_{k=1}^{K} \exp \eta_k + 1 \Rightarrow \psi = \sum_{k=1}^{K} \sum_{i=1}^{N} \gamma_{ik} = N$$

$$\Rightarrow \quad \pi_k = \frac{\sum_{i=1}^{N} \gamma_{ik}}{N}$$

Problem 1: basic GMM

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\gamma_{ik}\left[\log \pi_k + \frac{1}{2}\log\lambda_k - \frac{1}{2}(x_i-\mu_k)^2\lambda_k\right] \to \min_{\pi,\mu,\lambda}$$

Optimization w.r.t. μ_k:

$$\sum_{i=1}^{N}\gamma_{ik}(x_i-\mu_k)\lambda_k = 0 \quad \Rightarrow \quad \mu_k = \frac{\sum_{i=1}^{N}\gamma_{ik}x_i}{\sum_{i=1}^{N}\gamma_{ik}}$$

Problem 1: basic GMM

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\gamma_{ik}\left[\log\pi_k+\frac{1}{2}\log\lambda_k-\frac{1}{2}(x_i-\mu_k)^2\lambda_k\right]\to\min_{\pi,\mu,\lambda}$$

Optimization w.r.t. λ_k:

$$\sum_{i=1}^{N}\gamma_{ik}\left[\frac{1}{2\lambda_k}-\frac{1}{2}(x_i-\mu_k)^2\right]=0 \quad \Rightarrow \quad \lambda_k=\frac{\sum_{i=1}^{N}\gamma_{ik}}{\sum_{i=1}^{N}\gamma_{ik}(x_i-\mu_k)^2}$$

Problem 1: basic GMM

Resulting scheme:

- E-step: find posterior distribution for latent variables

$$p(Z|X) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\frac{\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \lambda_k^{-1})} \right]^{z_{ik}}$$

- M-step: find optimal values for parameters

$$\pi_k = \frac{\sum_{i=1}^{N} \gamma_{ik}}{N} \qquad \mu_k = \frac{\sum_{i=1}^{N} \gamma_{ik} x_i}{\sum_{i=1}^{N} \gamma_{ik}} \qquad \lambda_k = \frac{\sum_{i=1}^{N} \gamma_{ik}}{\sum_{i=1}^{N} \gamma_{ik} (x_i - \mu_k)^2}$$

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = p(\pi) \prod_{i=1}^{N} p(z_i|\pi) p(x_i|z_i, \mu, \lambda) =$$

$$= \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

How to train?

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = p(\pi)\prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) =$$

$$= \mathrm{Dir}(\pi|\alpha)\prod_{i=1}^{N}\prod_{k=1}^{K}\left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

How to train? EM-algorithm

- E-step: find posterior distribution for latent variables Z, π
- M-step: find optimal values for parameters μ, λ

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = p(\pi) \prod_{i=1}^{N} p(z_i|\pi) p(x_i|z_i, \mu, \lambda) =$$

$$= \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

How to train? EM-algorithm + mean field VI on E-step

- E-step: find posterior distribution for latent variables Z, π

$$p(Z, \pi|X, \mu, \lambda) \approx q(Z, \pi) = q(Z)q(\pi)$$

- M-step: find optimal values for parameters μ, λ

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = p(\pi) \prod_{i=1}^{N} p(z_i|\pi) p(x_i|z_i, \mu, \lambda) =$$
$$= \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are not conjugate. Check that there is a conditional conjugacy if we use the factorization $q(Z, \pi) = q(Z)q(\pi)$
- E-step: write down update rules for $q(Z)$ and $q(\pi)$.
- M-step: compute optimal values of μ, λ by maximizing $\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi|\mu, \lambda)$

Problem 2: + prior on π

Probabilistic model:
$$p(X, Z, \pi|\mu, \lambda) = p(\pi) \prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) =$$
$$= \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are not conjugate. Check that there is a conditional conjugacy if we use the factorization $q(Z, \pi) = q(Z)q(\pi)$
- E-step: write down update rules for $q(Z)$ and $q(\pi)$.
- M-step: compute optimal values of μ, λ by maximizing $\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi|\mu, \lambda)$

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi | \mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})\right]^{z_{ik}}$

Here different constants are denoted with the same letter C
for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Prior: $$p(Z, \pi) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{ik}} = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}}\right]$$

Here different constants are denoted with the same letter C
for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi|\mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$

Prior: $p(Z, \pi) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{ik}} = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}}\right]$

Posterior: $p(Z, \pi|X) \propto p(X, Z, \pi) = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} C^{z_{ik}}\right]$

Here different constants are denoted with the same letter C
for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi | \mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})\right]^{z_{ik}}$

Prior: $p(Z, \pi) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{ik}} = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}}\right]$

Posterior: $p(Z, \pi | X) \propto p(X, Z, \pi) = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} \boxed{C^{z_{ik}}}\right]$

no conjugacy

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi | \mu, \lambda) = \mathrm{Dir}(\pi | \alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right]^{z_{ik}}$

Prior: $p(Z, \pi) = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} \right]$

Posterior: $p(Z, \pi | X) \propto C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} C^{z_{ik}} \right]$

Here different constants are denoted with the same letter C
for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi | \mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})\right]^{z_{ik}}$

Fix Z

Prior: $p(Z, \pi) = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}}\right]$ $= C \prod_{k=1}^{K} \pi_k^C$

Posterior: $p(Z, \pi | X) \propto C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} C^{z_{ik}}\right]$ $= C \prod_{k=1}^{K} \pi_k^C$

conjugacy

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi | \mu, \lambda) = \mathrm{Dir}(\pi | \alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right]^{z_{ik}}$

Fix π

Prior: $p(Z, \pi) = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} \right]$ $\quad = C \prod_{k=1}^{K} \prod_{i=1}^{N} C^{z_{ik}}$

Posterior: $p(Z, \pi | X) \propto C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} C^{z_{ik}} \right]$ $\quad = C \prod_{k=1}^{K} \prod_{i=1}^{N} C^{z_{ik}}$

conjugacy

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Prior: $$p(Z, \pi) = C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}}\right]$$

Posterior: $$p(Z, \pi|X) \propto C \prod_{k=1}^{K} \left[\pi_k^C \prod_{i=1}^{N} \pi_k^{z_{ik}} C^{z_{ik}}\right]$$

Conditional conjugacy

$\downarrow$

$$q(Z, \pi) = q(Z)q(\pi)$$

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = p(\pi) \prod_{i=1}^{N} p(z_i|\pi) p(x_i|z_i, \mu, \lambda) =$$
$$= \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are not conjugate. Check that there is a conditional conjugacy if we use the factorization $q(Z, \pi) = q(Z) q(\pi)$
- E-step: write down update rules for $q(Z)$ and $q(\pi)$.
- M-step: compute optimal values of μ, λ by maximizing $\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi|\mu, \lambda)$

Problem 2: + prior on π

Probabilistic model:

$$p(X, Z, \pi|\mu, \lambda) = p(\pi) \prod_{i=1}^{N} p(z_i|\pi) p(x_i|z_i, \mu, \lambda) =$$

$$= \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Approximation: $p(Z, \pi|X, \mu, \lambda) \approx q(Z, \pi) = q(Z)q(\pi)$

Update rule for Variational inference:

$$\log q_j(\theta_j) = \mathbb{E}_{q_{i \neq j}} \log p(X, \theta) + Const, \quad \theta = (Z, \pi)$$

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi|\mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$

$$\begin{aligned}
\log q(Z) &= \mathbb{E}_{q(\pi)} \log p(X, Z, \pi) + Const = \\
&= \mathbb{E}_{q(\pi)} \left[\sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} (\log \pi_k + \log \mathcal{N}(x_i|\mu_k, \lambda_k^{-1}))\right] + Const = \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} (\mathbb{E}_{q(\pi)} \log \pi_k + \log \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})) + Const = \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \rho_{ik} + Const
\end{aligned}$$

$\Rightarrow$ $\boxed{q(Z) = \prod_{i=1}^{N} q(z_i), \quad q(z_i) = \frac{\prod_{k=1}^{K} {\rho_{ik}}^{z_{ik}}}{\sum_{k=1}^{K} \rho_{ik}}}$

Problem 2: + prior on π

Probabilistic model: $p(X, Z, \pi|\mu, \lambda) = \mathrm{Dir}(\pi|\alpha) \prod_{i=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$

$$\log q(\pi) = \mathbb{E}_{q(Z)} \log p(X, Z, \pi) + Const =$$

$$= \mathbb{E}_{q(Z)} \left[\sum_{k=1}^{K} (\alpha_k - 1) \log \pi_k + \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \pi_k \right] + Const =$$

$$= \sum_{k=1}^{K} (\alpha_k - 1) \log \pi_k + \sum_{i=1}^{N} \sum_{k=1}^{K} [\mathbb{E}_{q(Z)} z_{ik}] \log \pi_k + Const =$$

$$= \sum_{k=1}^{K} \log \pi_k \left(\alpha_k - 1 + \sum_{i=1}^{N} \mathbb{E}_{q(Z)} z_{ik} \right) + Const \quad \Rightarrow \quad \boxed{\begin{aligned} & q(\pi) = Dir(\pi|\alpha') \\ & {\alpha_k}' = \alpha_k + \sum_{i=1}^{N} \mathbb{E}_{q(Z)} z_{ik} \end{aligned}}$$

Problem 2: + prior on π

Probabilistic model: $$p(X, Z, \pi|\mu, \lambda) = p(\pi)\prod_{i=1}^{N} p(z_i|\pi)p(x_i|z_i, \mu, \lambda) =$$

$$= \mathrm{Dir}(\pi|\alpha)\prod_{i=1}^{N}\prod_{k=1}^{K}\left[\pi_k \mathcal{N}(x_i|\mu_k, \lambda_k^{-1})\right]^{z_{ik}}$$

Questions

- Check that likelihood and prior are not conjugate. Check that there is a conditional conjugacy if we use the factorization $q(Z, \pi) = q(Z)q(\pi)$
- E-step: write down update rules for $q(Z)$ and $q(\pi)$.
- M-step: compute optimal values of μ, λ by maximizing $\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi|\mu, \lambda)$

Problem 2: + prior on π

Rewrite $\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi | \mu, \lambda)$ as a function of μ, λ :

Problem 2: + prior on π

Rewrite $\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi | \mu, \lambda)$ as a function of μ, λ :

$$\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi) = \mathbb{E}_{q(Z,\pi)} \left[\log \mathrm{Dir}(\pi|\alpha) + \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left[\log \pi_k + \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right] \right] =$$

$$= \mathbb{E}_{q(Z,\pi)} \left[\sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right] + C =$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{E}_{q(Z)} z_{ik} \left[\frac{1}{2} \log \lambda_k - \frac{1}{2} (x_i - \mu_k)^2 \lambda_k \right] + C$$

Problem 2: + prior on π

Rewrite $\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi | \mu, \lambda)$ as a function of μ, λ :

$$\mathbb{E}_{q(Z,\pi)} \log p(X, Z, \pi) = \mathbb{E}_{q(Z,\pi)} \left[\log \mathrm{Dir}(\pi|\alpha) + \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left[\log \pi_k + \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right] \right] =$$

$$= \mathbb{E}_{q(Z,\pi)} \left[\sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right] + C =$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{E}_{q(Z)} z_{ik} \left[\frac{1}{2} \log \lambda_k - \frac{1}{2} (x_i - \mu_k)^2 \lambda_k \right] + C$$

→ Formulas for M-step are the same as earlier

Problem 2: + prior on π

Resulting scheme:

- E-step: mean-field approximation for posterior distribution of latent variables

$$q(Z) = \prod_{i=1}^{N} q(z_i), \quad q(z_i) = \prod_{k=1}^{K} {\rho_{ik}}^{z_{ik}} / \sum_{k=1}^{K} \rho_{ik}$$

$$q(\pi) = Dir(\pi|\alpha'), \quad {\alpha_k}' = \alpha_k + \sum_{i=1}^{N} \mathbb{E}_{q(Z)} z_{ik}$$

- M-step: find optimal values for parameters

$$\mu_k = \frac{\sum_{i=1}^{N} \mathbb{E}_{q(Z)} z_{ik} x_i}{\sum_{i=1}^{N} \mathbb{E}_{q(Z)} z_{ik}} \qquad \lambda_k = \frac{\sum_{i=1}^{N} \mathbb{E}_{q(Z)} z_{ik}}{\sum_{i=1}^{N} \mathbb{E}_{q(Z)} z_{ik} (x_i - \mu_k)^2}$$