

Approximate Bayesian inference

Ekaterina Lobacheva

Research Fellow at Samsung-HSE Laboratory



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

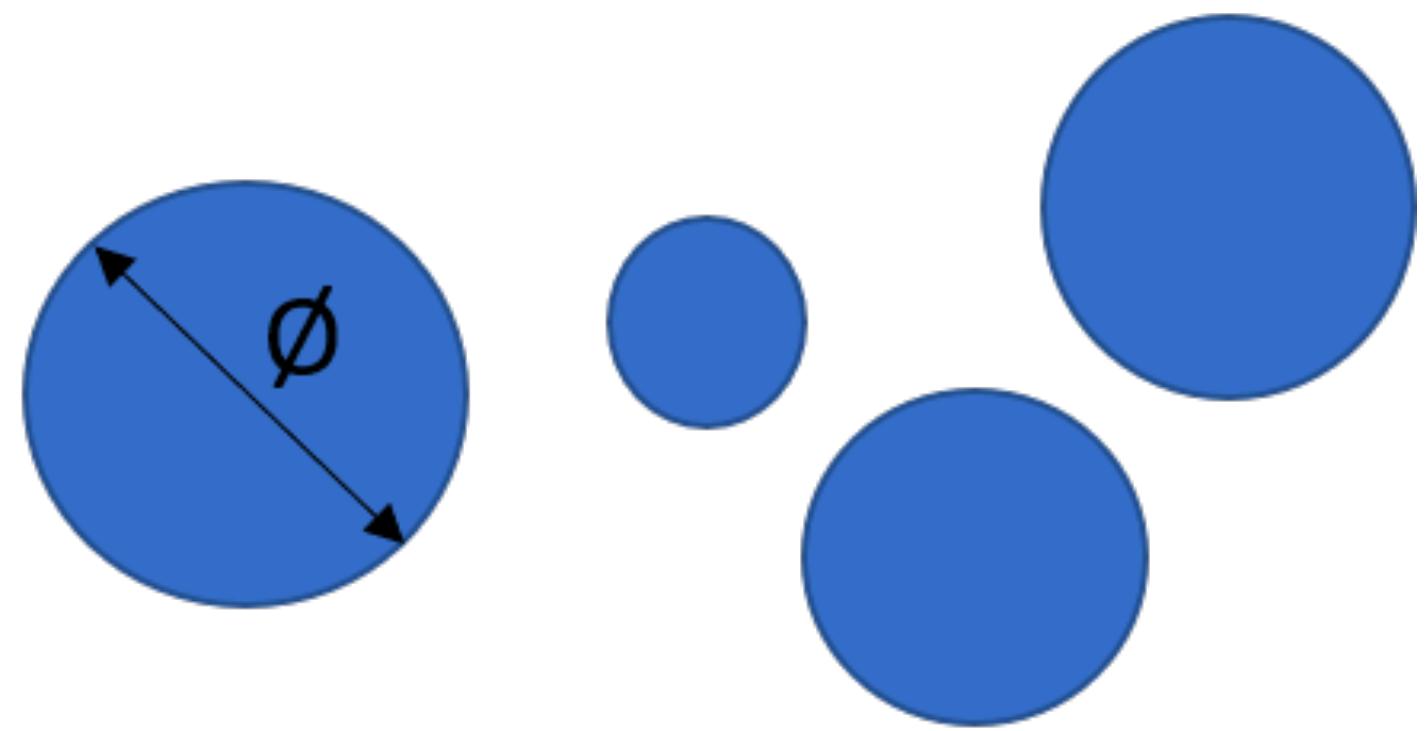
Gaussian mixture model:

- K Gaussian components with probabilities $\pi = (\pi_1, \dots, \pi_K)$
- 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



GMM



- Hard clustering
- Diagonal covariance matrices

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

Gaussian mixture model

Basic probabilistic model:

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

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:

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

How to train?

Problem 1: basic GMM

Probabilistic model:

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

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:

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

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:

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

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:

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

Problem 1: Gaussian mixture model

Probabilistic model:

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

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

Problem 1: Gaussian mixture model

Probabilistic model:

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

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

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

Problem 1: Gaussian mixture model

Probabilistic model:

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

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

$$= \boxed{\prod_{i=k}^K \prod_{i=1}^N C^{z_{ik}}}$$

conjugate

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

$$= \boxed{\prod_{i=k}^K \prod_{i=1}^N C'^{z_{ik}}}$$

Problem 1: basic GMM

Probabilistic model:

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

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 \rightarrow posterior is Categorical

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

Problem 1: Bayesian framework

Likelihood and prior are conjugate \rightarrow posterior is Categorical

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

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

Problem 1: Bayesian framework

Likelihood and prior are conjugate → posterior is Categorical

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

Normalization

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

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

Problem 1: Bayesian framework

Likelihood and prior are conjugate \rightarrow posterior is Categorical

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

Normalization

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

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

Problem 1: Bayesian framework

Likelihood and prior are conjugate \rightarrow posterior is Categorical

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

Normalization

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

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

Result:

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

Problem 1: basic GMM

Probabilistic model:

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

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 :

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

Problem 1: basic GMM

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

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

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

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

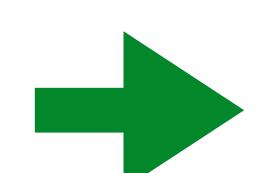
Problem 1: basic GMM

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

Differentiation:

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

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


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

Problem 1: basic GMM

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

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

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

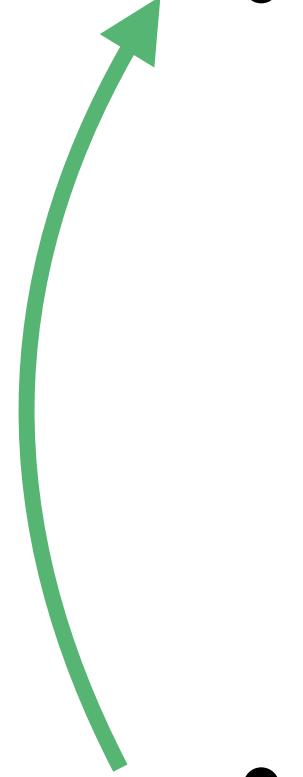
Optimization w.r.t. λ_k :

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

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

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

Problem 2: + prior on π

Probabilistic model:

$$\begin{aligned} p(X, Z, \pi | \mu, \lambda) &= p(\pi) \prod_{i=1}^N p(z_i | \pi) p(x_i | z_i, \mu, \lambda) = \\ &= \text{Dir}(\pi | \alpha) \prod_{i=1}^N \prod_{k=1}^K [\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})]^{z_{ik}} \end{aligned}$$

How to train?

Problem 2: + prior on π

Probabilistic model:

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

How to train? EM-algorithm

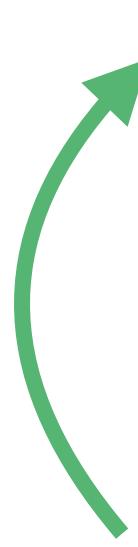
- 
- E-step: find posterior distribution for latent variables Z, π
 - M-step: find optimal values for parameters μ, λ

Problem 2: + prior on π

Probabilistic model:

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

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:

$$\begin{aligned} p(X, Z, \pi | \mu, \lambda) &= p(\pi) \prod_{i=1}^N p(z_i | \pi) p(x_i | z_i, \mu, \lambda) = \\ &= \text{Dir}(\pi | \alpha) \prod_{i=1}^N \prod_{k=1}^K [\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})]^{z_{ik}} \end{aligned}$$

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 $\mathbb{E}_{p(Z, \pi | X, \mu, \lambda)} \log p(X, Z, \pi | \mu, \lambda)$

Problem 2: + prior on π

Probabilistic model:

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

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 $\mathbb{E}_{p(Z, \pi | X, \mu, \lambda)} \log p(X, Z, \pi | \mu, \lambda)$

Problem 2: + prior on π

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

Problem 2: + prior on π

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

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

Problem 2: + prior on π

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

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

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

Problem 2: + prior on π

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

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

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

no conjugacy

Problem 2: + prior on π

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

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

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

Problem 2: + prior on π

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

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

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

Fix Z

$$= C \prod_{k=1}^K \pi_k^\psi$$

$$= C'' \prod_{k=1}^K \pi_k^{\psi'}$$

conjugacy

Problem 2: + prior on π

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

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

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

Fix π

$$= C \prod_{k=1}^K \prod_{i=1}^N \xi^{z_{ik}}$$

$$= C'' \prod_{k=1}^K \prod_{i=1}^N \xi'^{z_{ik}}$$

conjugacy

Problem 2: + prior on π

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

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

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

Conditional conjugacy



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

Problem 2: + prior on π

Probabilistic model:

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

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 $\mathbb{E}_{p(Z, \pi | X, \mu, \lambda)} \log p(X, Z, \pi | \mu, \lambda)$

Problem 2: + prior on π

Probabilistic model:

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

Approximation:

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

Update rule for Variational inference:

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

Problem 2: + prior on π

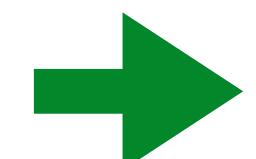
Probabilistic model: $p(X, Z, \pi | \mu, \lambda) = \text{Dir}(\pi | \alpha) \prod_{i=1}^N \prod_{k=1}^K [\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})]^{z_{ik}}$

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

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

$$= \sum_{i=1}^N \sum_{k=1}^K z_{ik} (\mathbb{E}_{q(\pi)} \log \pi_k + \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})) + Const =$$

$$= \sum_{i=1}^N \sum_{k=1}^K z_{ik} \log \rho_{ik} + Const$$



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

Problem 2: + prior on π

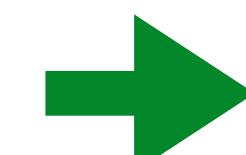
Probabilistic model: $p(X, Z, \pi | \mu, \lambda) = \text{Dir}(\pi | \alpha) \prod_{i=1}^N \prod_{k=1}^K [\pi_k \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})]^{z_{ik}}$

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

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

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

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



$$q(\pi) = \text{Dir}(\pi | \alpha')$$

$$\alpha_k' = \alpha_k + \sum_{i=1}^N \mathbb{E}_{q(Z)} z_{ik}$$

Problem 2: + prior on π

Probabilistic model:

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

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
 $\mathbb{E}_{p(Z, \pi | X, \mu, \lambda)} \log p(X, Z, \pi | \mu, \lambda)$

Problem 2: + prior on π

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

Problem 2: + prior on π

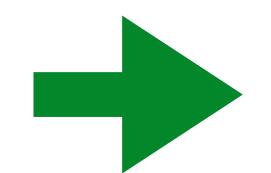
Rewrite $\mathbb{E}_{p(Z, \pi | X)} \log p(X, Z, \pi)$ as a function of μ, λ :

$$\begin{aligned}\mathbb{E}_{p(Z, \pi | X)} \log p(X, Z, \pi) &= \mathbb{E}_{p(Z, \pi | X)} \left[\log \text{Dir}(\pi | \alpha) + \sum_{i=1}^N \sum_{k=1}^K z_{ik} [\log \pi_k + \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})] \right] = \\ &= \mathbb{E}_{p(Z, \pi | X)} \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik} \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right] + C = \\ &= \left\{ \mathbb{E}_{p(Z | X)} z_{ik} = p(z_{ik} = 1 | X) = \gamma_{ik} \right\} = \\ &= \sum_{i=1}^N \sum_{k=1}^K \gamma_{ik} \left[\frac{1}{2} \log \lambda_k - \frac{1}{2} (x_i - \mu_k)^2 \lambda_k \right] + C\end{aligned}$$

Problem 2: + prior on π

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

$$\begin{aligned}\mathbb{E}_{p(Z, \pi | X)} \log p(X, Z, \pi) &= \mathbb{E}_{p(Z, \pi | X)} \left[\log \text{Dir}(\pi | \alpha) + \sum_{i=1}^N \sum_{k=1}^K z_{ik} [\log \pi_k + \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1})] \right] = \\ &= \mathbb{E}_{p(Z, \pi | X)} \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik} \log \mathcal{N}(x_i | \mu_k, \lambda_k^{-1}) \right] + C = \\ &= \left\{ \mathbb{E}_{p(Z | X)} z_{ik} = p(z_{ik} = 1 | X) = \gamma_{ik} \right\} = \\ &= \sum_{i=1}^N \sum_{k=1}^K \gamma_{ik} \left[\frac{1}{2} \log \lambda_k - \frac{1}{2} (x_i - \mu_k)^2 \lambda_k \right] + C\end{aligned}$$

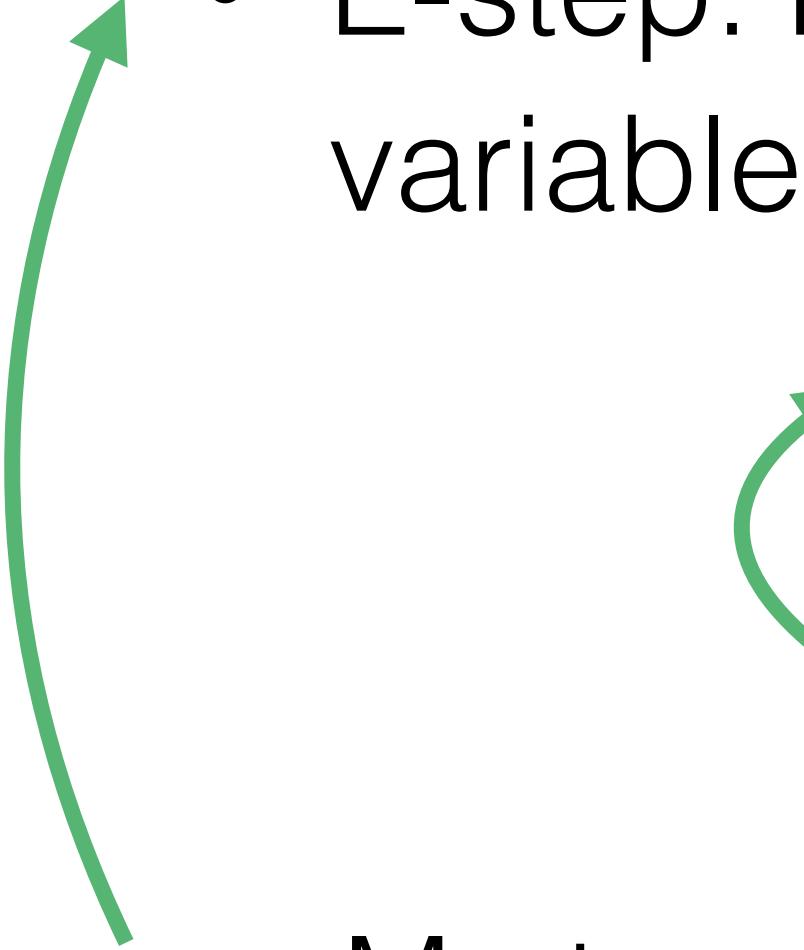


Formulas for M-step
are the same as earlier

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

$$\mu_k = \frac{\sum_{i=1}^N \gamma_{ik} x_i}{\sum_{i=1}^N \gamma_{ik}}$$

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