A Review of Variation Bayesian Gaussian Mixture Model

資工21 106033233 周聖諺

Introduction

The disadvantage of K-Means

- Hard to decide the number of clusters
- Poor performance on unbalanced dataset

Variational Bayesian Expectation Maximization(VBEM)

- Only need to choose the maximum number of clusters, self-adapt to the best
- Better performance on unbalanced dataset

EM In General Form

In naive EM, the goal is to maximize likelihood $\mathcal{L}(Y; \theta)$

$$rg \max_{ heta} \mathcal{L}(Y; heta) = rg \max_{ heta} \ log \int_{Z} p(Y, Z; heta) dZ$$

where Y is observing data, Z is latent variable, heta is the parameter

With ELBO, we can derive the lower bound $\mathcal{L}(\theta,\gamma)$ of likelihood $\mathcal{L}(Y;\theta)$

$$egin{aligned} \mathcal{L}(heta,\gamma) &= \mathbb{E}_q[log(rac{p(Y,Z; heta)}{q(Z;\gamma)})] \ &= \int_Z q(Z;\gamma)log \, rac{p(Y,Z; heta)}{q(Z;\gamma)} dZ \ &= log \, p(Y; heta) - KL[q(Z;\gamma)||p(Z|Y)] \ &= \mathcal{L}(Y; heta) - KL[q(Z;\gamma)||p(Z|Y)] \end{aligned}$$

EM In General Form

Thus

$$\mathcal{L}(heta,\gamma) = \mathcal{L}(Y; heta) - KL[q(Z;\gamma)||p(Z|Y)]$$

Since the KL-divergence always ≥ 0

$$rg \max_{ heta} \mathcal{L}(Y; heta) \geq rg \max_{ heta, \gamma} \mathcal{L}(heta, \gamma)$$

With KKT and Lagrange multiplier, the optimization problem can be written as

$$rg \max_{ heta, \gamma} \mathcal{L}(heta, \gamma) = rg \max_{ heta, \gamma} \mathcal{L}(Y; heta) - eta KL[q(Z; \gamma) || p(Z|Y)]$$

EM In General Form

Pseudo Code

Iterate until heta converge

• E Step at k-th iteration

$$\gamma_{k+1} = rg \max_{\gamma_k} \mathcal{L}(heta_k, \gamma_k)$$

• M Step at k-th iteration

$$heta_{k+1} = rg \max_{ heta_k} \mathcal{L}(heta_k, \gamma_{k+1})$$

Variational Bayesian Expectation Maximization(VBEM)

In VBEM, we consider an additional prior $p(\theta; \lambda)$

$$egin{aligned} log \ p(Y) &= log \int_{Z, heta} p(Y,Z, heta;\lambda) dZ d heta \ &= log \ \mathbb{E}_{q(Z;\phi^Z)q(heta;\phi^ heta)} [rac{p(Y,Z| heta)p(heta;\lambda)}{q(Z;\phi^Z)q(heta;\phi^ heta)}] \ &\geq \mathbb{E}_{q(Z;\phi^Z)q(heta;\phi^ heta)} [log \ rac{p(Y,Z| heta)p(heta;\lambda)}{q(Z;\phi^Z)q(heta;\phi^ heta)}] \end{aligned}$$

Thus, we get the ELBO $\mathcal{L}(\phi^Z,\phi^ heta)$

$$\mathcal{L}(\phi^Z,\phi^ heta) = \mathbb{E}_{q(Z;\phi^Z)q(heta;\phi^ heta)}[log \; rac{p(Y,Z| heta)p(heta;\lambda)}{q(Z;\phi^Z)q(heta;\phi^ heta)}]$$

Variational Bayesian Expectation Maximization(VBEM)

According to the general form of EM

$$rg \max_{\gamma_k} \mathcal{L}(heta_k, \gamma_k) \ rg \max_{ heta_k} \mathcal{L}(heta_k, \gamma_k)$$

We can derive

$$egin{aligned} rac{d}{d\phi^Z} \mathcal{L}(\phi^Z,\phi^ heta) &= 0, \quad ln \ q(Z;\phi^Z) \propto \mathbb{E}_{q(heta;\phi^ heta)}[log \ p(Y,Z, heta)] \ rac{d}{d\phi^ heta} \mathcal{L}(\phi^Z,\phi^ heta) &= 0, \quad ln \ q(heta;\phi^ heta) \propto \mathbb{E}_{q(Z;\phi^Z)}[log \ p(Y,Z, heta)] \end{aligned}$$

Variational Bayesian Expectation Maximization(VBEM)

Pseudo Code

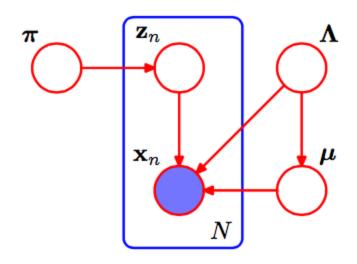
Iterate until $\mathcal{L}(\phi^Z,\phi^ heta)$ converge

ullet E Step: Update the variational distribution on Z $q(Z;\phi^Z) \propto e^{(\mathbb{E}_{q(heta;\phi^ heta)}[log\;p(Y,Z, heta)])}$

ullet M Step: Update the variational distribution on heta

$$q(heta;\phi^{ heta}) \propto e^{(\mathbb{E}_{q(Z;\phi^Z)}[log\;p(Y,Z, heta)])}$$

Variational Bayesian Gaussian Mixture Model(VB-GMM)



$$p(X,Z,\pi,\mu,\Lambda) = p(X|Z,\pi,\mu,\Lambda)p(Z|\pi)p(\pi)p(\mu|\Lambda)p(\Lambda)$$

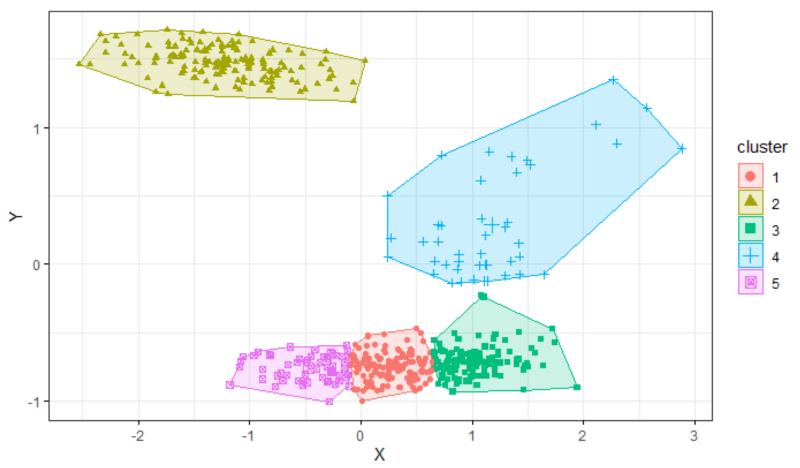
- ullet $p(X|Z,\pi,\mu,\Lambda)$ denotes the Gaussian Mixture Model
- $p(Z|\pi)$ denotes the Latent Variables
- ullet $p(\pi)$ denotes the **Dirichlet Prior Distribution Over The Latent Variables** Z
- ullet $p(\mu|\Lambda)p(\Lambda)$ denotes the Gaussian-Wishart Prior Distribution Over GMM X

Ignore The Complex Derivation

K-means Cluster Plot

Fit On 5 Modal Simulation Dataset Generated by GMM

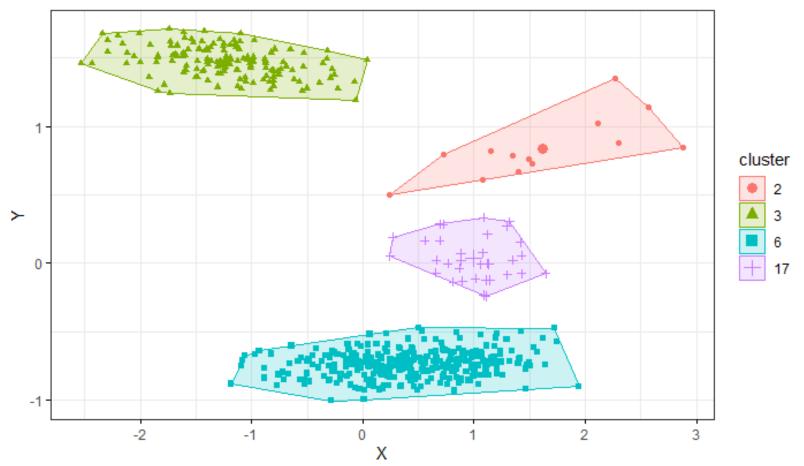
Simulation Cluster Plot of K-Means



VB-GMM Cluster Plot

Fit On 5 Modal Simulation Dataset Generated by GMM

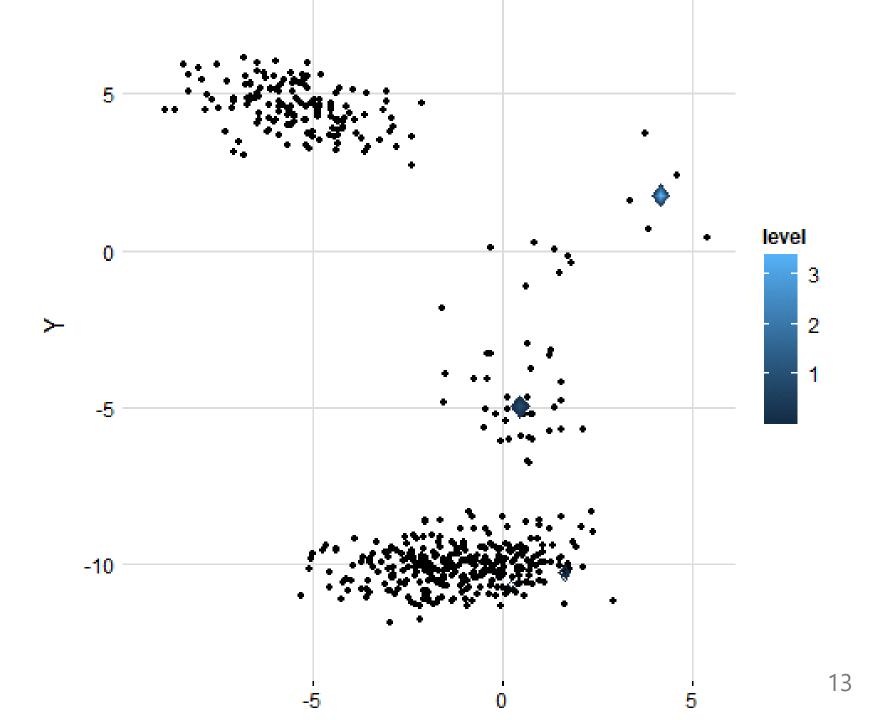




VB-GMM Contour Plot Animation

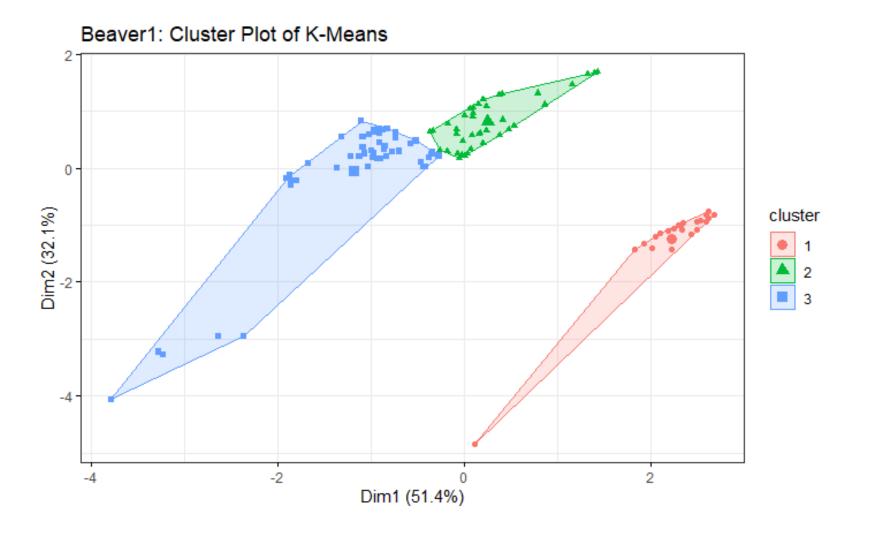
Fit On 5 Modal Simulation dataset

Link



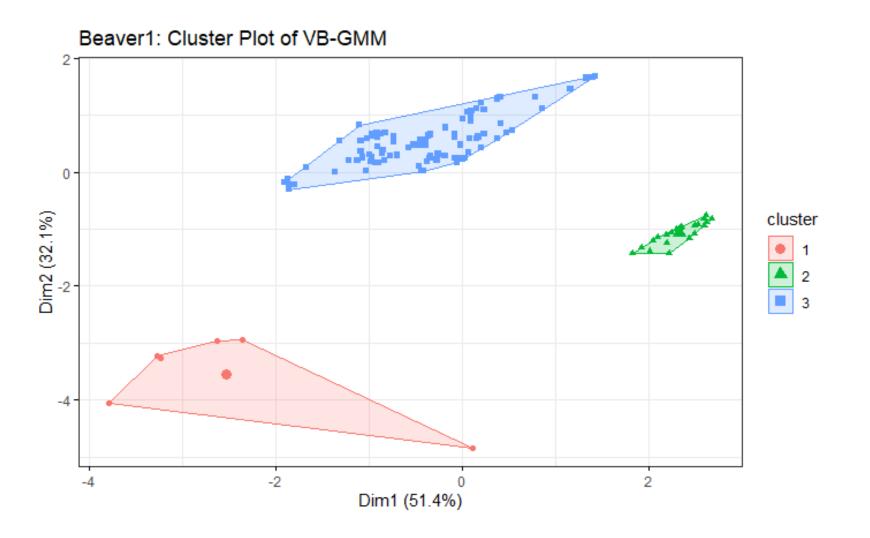
K-means Cluster Plot

Fit On Beaver1 dataset



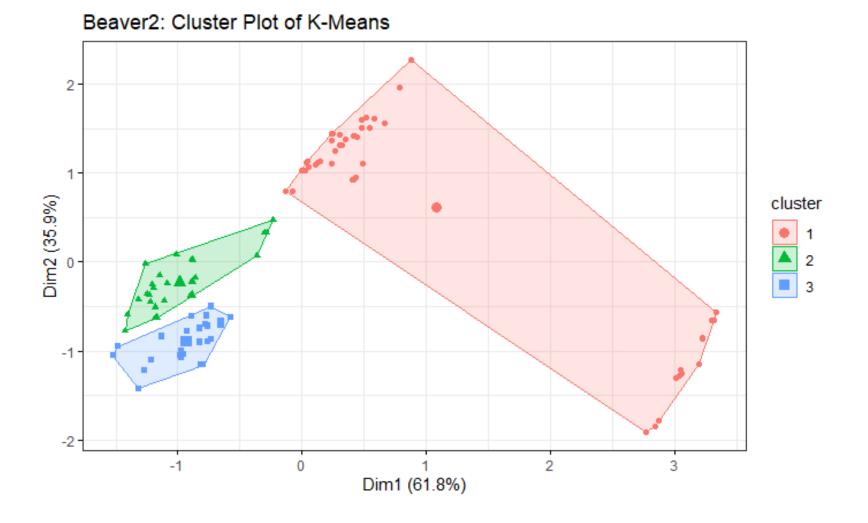
VB-GMM Cluster Plot

Fit On Beaver1 dataset



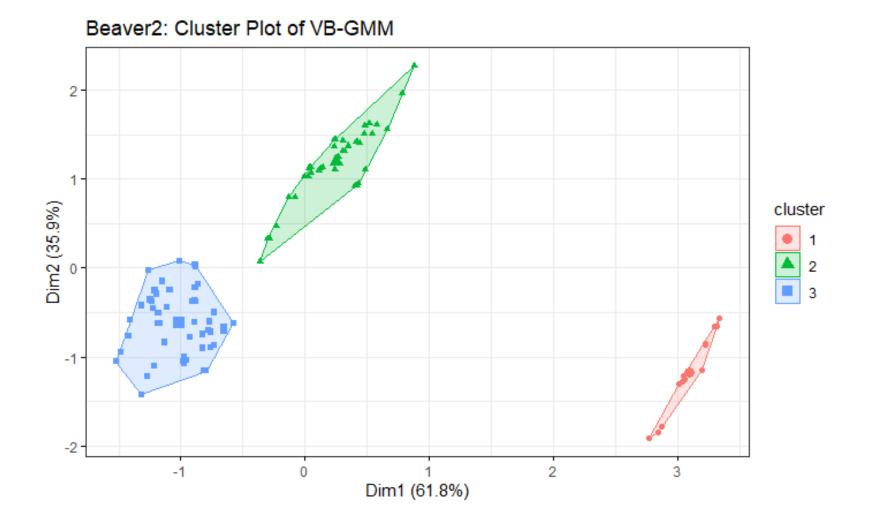
K-means Cluster Plot

Fit On Beaver2 dataset



VB-GMM Cluster Plot

Fit On Beaver2 dataset



Thanks For Listening