变分贝叶斯

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

1. 参数估计

- 参数估计
  - ✓ 贝叶斯估计

$$p(\theta|X)=\frac{p(X|\theta)\,\mathrm{p}(\theta)}{p(X)}=\frac{p(X|\theta)\,\mathrm{p}(\theta)}{\int p(X,\theta)\,\mathrm{d}\,\theta}\propto p(X|\theta)\,\mathrm{p}(\theta)$$

  - ✓ 最大后验估计

$$p(X|\theta)\,\mathrm{p}(\theta)$$

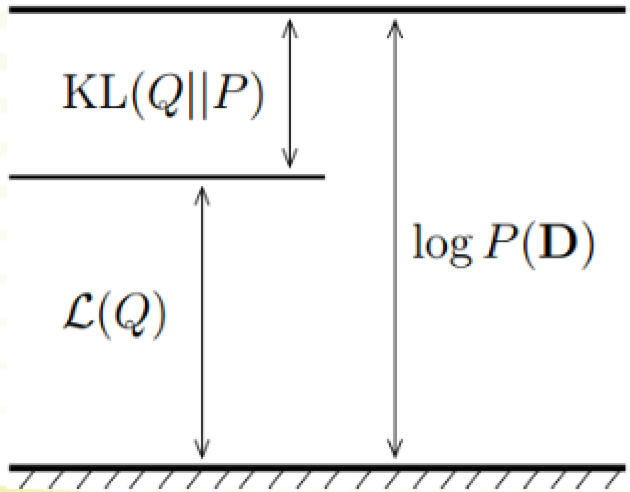
  - ✓ 极大似然估计

$$p(X|\theta)$$

2. 变分贝叶斯

- 变分贝叶斯
  - ✓ 贝叶斯估计中分母P(X)往往很难求，于是找一个简单的函数q(θ)来近似P(θ|X)，即q(θ) ≈ P(θ|X)
  - ✓ 如何评价q(θ)与P(θ|X)之间的近似程度？——Kullback-Leibler散度
  - ✓ 目标函数：min KL(q(θ)|| P(θ|X))
  - ✓ ln P(X)=L(q)+KL(q||p)，而ln P(X)是与θ无关的常量，不变
  - ✓ min KL(q(θ)|| P(θ|X))⇔max L(q)

$$L(q)=\int q(\theta)\ln\frac{p(X,\theta)}{q(\theta)}\mathrm{d}\,z$$



- 变分贝叶斯
  - ✓ 平均场理论
  - ✓ 根据平均场理论，变分分布q(θ)可以因式分解为K个互不相交的部分

$$q(\theta) = \prod_{k=1}^{K} q_k(\theta_k)$$

  - ✓ 求解得： $\ln q_j^*(\theta_j) = \mathrm{E}_{i \neq j}(\ln p(X,\theta)) + \mathrm{const}$
  - ✓ 两边同时取指数，再归一化，得

$$q_j^*(\theta_j) = \frac{\exp(\mathrm{E}_{i \neq j}(\ln p(X,\theta)))}{\int \exp(\mathrm{E}_{i \neq j}(\ln p(X,\theta)))\,\mathrm{d}z_j}$$

3. Student's Distribution Mixture Model

➢ Student's Distribution Mixture Model

$$f(\mathrm{x} \mid \theta_S) = \sum_{k=1}^{K} \pi_k S(x; \mu_k, \Lambda_k, v_k)$$

$$S(x \mid \mu, \Lambda, v) = \frac{\Gamma(\frac{v+D}{2})}{\Gamma(\frac{v}{2})} \frac{|\Lambda|^{1/2}}{(v\pi)^{1/2}} [1 + \frac{1}{v}(x-\mu)^{\mathrm{T}} \Lambda (x-\mu)]^{-\frac{v+D}{2}}$$

$$= \int_0^{+\infty} \mathcal{N}(\mathrm{x} \mid \mu, u\Lambda) \mathcal{G}(u \mid \frac{v}{2}, \frac{v}{2}) \mathrm{d}u$$

t分布可表示为具有相同均值、不同尺度精度的高斯分布的无限混合。

其中，π为混合比例，v：自由度，μ：均值，Λ：精度矩阵。u：尺度因子

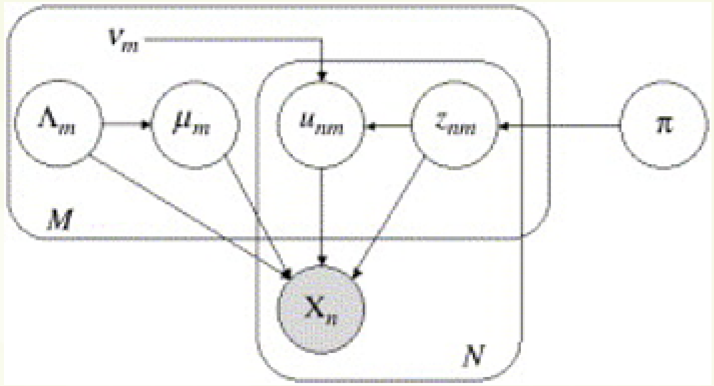
➢ Student's Distribution Mixture Model

先验分布： π~Dirichlet分布，(μ, Λ)~Gaussian-Wishart分布

$$p(z_n \mid \theta_S) = \prod_{k=1}^{K} \pi_k^{z_{nk}}$$

$$p(u_n \mid z_n, \theta_S) = \prod_{k=1}^{K} \mathcal{G}\left(u_{nk} \mid \frac{v_k}{2}, \frac{v_k}{2}\right)^{z_{nk}}$$

$$p(x_n \mid u_n, z_n, \theta_S) = \prod_{k=1}^{K} N(x_n \mid \mu_k, u_{nk}\Lambda_k)^{z_{nk}}$$



$$p(X, U, Z, \theta_S) = p(x_n \mid u_n, z_n, \theta_S)\, p(u_n \mid z_n, \theta_S)\, p(z_n \mid \theta_S)\, p(\pi)\, \mathrm{p}(\mu, \Lambda)$$

➢ VBE-step: $q(u_n, z_n) \propto \exp(\mathrm{E}_{\theta_S}\{\log p(x_n, u_n, z_n, \theta_S)\})$

➢ VBM-step: $q(\theta_S) \propto \exp(\mathrm{E}_{U,Z}\{\log p(x_n, u_n, z_n, \theta_S)\})$

4. 参考文献

[1] Cédric Archambeau, Verleysen M . Robust Bayesian clustering[J]. Neural Networks, 2007, 20(1):129-138.

[2] 详细请参考：华俊豪博客-变分推理、变分贝叶斯算法理解与推导

[3] David M. Blei：变分推断讲义

[4] Corduneanu, A., & Bishop, C. (2001). Variational Bayesian model selection for mixture distributions. In *Artifical Intelligence and Statistics* (pp. 27–34)