EECS E6892 Bayesian Models for Machine Learning
Homework 4

John Min

April 18, 2014

1 Normal-Wishart prior

The multivariate analog ofthe normal-gamma prior is the normal-Wishart prior.

We are given observations $x_1, \dots, x_n$ from a d-dimensional multivariate Gaussian with a Normal-Wishart prior on the mean and precision matrix: $x_i \sim N(\mu, \Lambda^{-1}), \mu|\Lambda \sim N(m, (\frac{1}{a}\Lambda)^{-1}), \Lambda \sim Wishart(\nu, B)$.

1.1 Posterior $p(\mu, \Lambda | x_1, \dots, x_n)$

$$\begin{aligned} p(\mu, \Lambda | x_1, \dots, x_N) = & \\ & \sim N\Big(\mu \;\Big|\; \frac{an\bar{x} + m}{an+1}, \big[(n + \frac{1}{a})\Lambda\big]^{-1}\Big) \cdot W(B_n, \nu_n) \\ & \sim NW(\mu_n, \Lambda_n, B_n, \nu_n) \end{aligned}$$

where $\mu_n = \frac{an\bar{x}+m}{an+1}$, $\Lambda_n = \big[(n+\frac{1}{a})\Gamma\big]^{-1}$, $B_n = \Big[B^{-1} + \sum_i^N (x_i - \bar{x})(x_i - \bar{x})^\top + \dfrac{n}{an+1}(\bar{x} - m)(\bar{x} - m)^\top\Big]^{-1}$, $\nu_n = \nu + n$.

1.2 Marginal likelihood of a singular data point

For a single vector, $p(x) = \int_\Lambda \int_\mu p(x|\mu, \Lambda) p(\mu|\Lambda) p(\Lambda) \partial\mu \partial\Lambda$.

Given this integral, we can integrate out a Normal-Wishart distribution and thus, we are left with a normalization ratio.

$$p(x) = \pi^{-d/2} \frac{\Gamma}{} \Big(\Big)^{d/2}$$

2 Expectation Maximization for Gaussian Mixture Models

2.1 The EM algorithm

1. initialize μ_k, covariances Σ_k, mixing coefficients π_k, and evaluate the initia value of the log-likelihood
2. **(E-step)**: evaluate responsibilities

   $\gamma(z_{nk}) \leftarrow \pi_k \cdot N(x_n|\mu_k, \Sigma_k) / \sum_{j=1}^K \pi_j N(x_n|\mu_j, \Sigma_j)$

3. **(M-step)**: re-estimate parameters using current responsibilities

   $\mu_k^{\text{new}} \leftarrow \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n$

   $\Sigma_k^{\text{new}} \leftarrow \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \Big(x_n - \mu_k^{\text{new}}\Big)\Big(x_n - \mu_k^{\text{new}}\Big)^\top$

   $\pi_k^{\text{new}} \leftarrow \frac{N_k}{N}$ where $N_k = \sum_{n=1}^N \gamma(z_{nk})$

4. Evaluate log-likelihood
$$\ln p(x|\mu, \Sigma, \pi) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k \right\}$$

2.2 Plots of log-likelihood

2.2.1 K=2

2.2.2 K=4

2.2.3 K=6

2.2.4 K=8

2.2.5 K=10

3 Dirichlet Process Gaussian Mixture Model (D.P.G.M.M.)

3.1 Normal-Wishart prior