

PRML の 9 章の数式の補足

サイボウズ・ラボ 光成滋生

2011 年 7 月 22 日

1 概要

この文章は『パターン認識と機械学習』(以下 PRML) の 9 章の式変形を一部埋めたものです. 間違い, 質問などございましたら herumi@nifty.com または [twitterID:herumi](https://twitter.com/herumi) までご連絡ください. 面倒なので特に紛らわしいと思わない限り x を x と書いたりします.

2 復習

よく使ういくつかの式を書いておく. どれも今までに既に示したものである.

<https://github.com/herumi/prml/raw/master/prml2.pdf>,

<https://github.com/herumi/prml/raw/master/prml3.pdf> を参照.

行列について

$$\begin{aligned}x^T A x &= \text{tr}(A x x^T) \\ \frac{\partial}{\partial A} \log |A| &= (A^{-1})^T \\ \frac{\partial}{\partial x} \log |A| &= \text{tr}(A^{-1} \frac{\partial}{\partial x} A) \\ \frac{\partial}{\partial A} \text{tr}(A^{-1} B) &= -(A^{-1} B A^{-1})^T\end{aligned}$$

3 混合ガウス分布

離散的な潜在変数を用いた混合ガウス分布の定式化. K 次元 2 値確率変数 z を考える (どれか一つの成分のみが 1 であとは 0). つまり

$$\sum_k z_k = 1.$$

z の種類は K 個である. $0 \leq \pi_k \leq 1$ という係数を用いて

$$p(z_k = 1) = \pi_k$$

という確率分布を与える.

$$p(z) = \prod_k \pi_k^{z_k}.$$

$$p(x|z_k = 1) = \mathcal{N}(x|\mu_k, \Sigma_k)$$

なので

$$P(x|z) = \prod_k \mathcal{N}(x|\mu_k, \Sigma_k)^{z_k}.$$

これらを合わせて

$$\begin{aligned} p(x) &= \sum_z p(z)p(x|z) \\ &= \sum_z \prod_k (\pi_k \mathcal{N}(x|\mu_k, \Sigma_k))^{z_k} \\ &\quad z_k \text{ はどれか一つのみが } 1 \text{ (そのとき } \pi_k \text{) であとは } 0 \text{ なので} \\ &= \sum_k \pi_k \mathcal{N}(x|\mu_k, \Sigma_k). \end{aligned}$$

x が与えられたときの z の条件付き確率 $p(z_k = 1|x)$ を $\gamma(z_k)$ とする.

$$\gamma(z_k) = \frac{p(z_k = 1)p(x|z_k = 1)}{\sum_j p(z_j = 1)p(x|z_j = 1)} = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}.$$

これを混合要素 k が観測値 x に対する負担率という.

4 混合ガウス分布の EM アルゴリズム

混合ガウス分布において観測したデータ集合を $X^T = \{x_1, \dots, x_N\}$, 対応する潜在変数を $Z^T = \{z_1, \dots, z_N\}$ とする. X は $N \times D$ 行列で Z は $N \times K$ 行列.

対数尤度関数の最大点の条件をもとめる.

$$F = \log p(X|\pi, \mu, \Sigma) = \sum_{n=1}^N \log \left(\sum_{j=1}^K \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j) \right)$$

とする.

$$\frac{\partial}{\partial \mu} \log \mathcal{N}(x|\mu, \Sigma) = \frac{\partial}{\partial \mu} \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) = \Sigma^{-1} (x - \mu)$$

を使うと

$$\begin{aligned} \frac{\partial}{\partial \mu_k} F &= \sum_n \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \Sigma_k^{-1} (x_n - \mu_k) \\ &= \Sigma_k^{-1} \left(\sum_n \gamma(z_{nk}) (x_n - \mu_k) \right) = 0. \end{aligned}$$

よって

$$\begin{aligned} \sum_n \gamma(z_{nk}) x_n - \left(\sum_n \gamma(z_{nk}) \right) \mu_k &= 0. \\ N_k &= \sum_n \gamma(z_{nk}) \end{aligned}$$

とおくと

$$\mu_k = \frac{1}{N_k} \sum_n \gamma(z_{nk}) x_n.$$

これは μ_k が X の重みつき平均であることを示している.

次に Σ_k に関する微分を考える.

$$\mathcal{N} = \mathcal{N}(x|\mu, \Sigma)$$

のとき

$$\log \mathcal{N} = -\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \text{tr}(\Sigma^{-1}(x - \mu)(x - \mu)^T)$$

なので

$$\frac{\partial}{\partial \Sigma}(\log \mathcal{N}) = -\frac{1}{2}(\Sigma^{-1})^T + \frac{1}{2}(\Sigma^{-1}(x - \mu)(x - \mu)^T \Sigma^{-1})^T.$$

よって

$$\begin{aligned} \frac{\partial}{\partial \Sigma_k} F &= \sum_n \gamma(z_{nk}) \frac{\partial}{\partial \Sigma_k} \mathcal{N}(x_n|\mu_k, \Sigma_k) \\ &= \sum_n \gamma(z_{nk}) \left(-\frac{1}{2}(\Sigma_k^{-1}) + \frac{1}{2}(\Sigma_k^{-1}(x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1})^T \right) = 0. \end{aligned}$$

よって

$$\sum_n \gamma(z_{nk}) (I - (x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1}) = 0.$$

$$\Sigma_k = \frac{1}{N_k} \sum_n \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T.$$

最後に π_k に関する微分を考える. $\sum_k \pi_k = 1$ の制約を入れる.

$$G = F + \lambda \left(\sum_k \pi_k - 1 \right)$$

とすると

$$\frac{\partial}{\partial \pi_k} F = \sum_n \frac{\mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} + \lambda = 0.$$

$\mathcal{N}_{nk} = \mathcal{N}(x_n|\mu_k, \Sigma_k)$ とおくと

$$N = \sum_k N_k = \sum_{n,k} \gamma(z_{nk}) = \sum_k \pi_k \left(\sum_n \frac{\mathcal{N}_{nk}}{\sum_j \pi_j \mathcal{N}_{nj}} \right) = - \sum_k \pi_k \lambda = -\lambda.$$

よって

$$\begin{aligned} \sum_n \frac{\pi_k \mathcal{N}_{nk}}{\sum_j \pi_j \mathcal{N}_{nj}} &= \pi_k N. \\ \pi_k &= \frac{1}{N} \sum_n \gamma(z_{nk}) = \frac{N_k}{N}. \end{aligned}$$

5 K -means との関連

式 (9.43) は不正確. E ではなく ϵE を考えないと (9.43) の右辺にはならない. 式 (9.40) を E とおく.

$$E = \sum_{n,k} \gamma(z_{nk}) (\log \pi_k + \log \mathcal{N}(x_n|\mu_k, \Sigma_k)).$$

ϵE に

$$\mathcal{N}(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi\epsilon)^{D/2}} \exp\left(-\frac{1}{2\epsilon} \|x - \mu_k\|^2\right)$$

を代入する.

$$\epsilon E = \sum_{n,k} \gamma(z_{nk}) \left(\epsilon \log \pi_k - \frac{D}{2} \epsilon \log(2\pi\epsilon) - \frac{1}{2} \|x_n - \mu_k\|^2 \right).$$

$\epsilon \rightarrow 0$ で

$$\gamma(z_{nk}) \rightarrow r_{nk}.$$

$$\epsilon \log \pi_k \rightarrow 0.$$

$$\epsilon \log(2\pi\epsilon) \rightarrow 0$$

より

$$\epsilon E \rightarrow -\frac{1}{2} \sum_{n,k} r_{nk} \|x_n - \mu_k\|^2 = -J.$$

よって期待完全データ対数尤度の最大化は J の最小化と同等.