10.1.3 変分推論-ガウス分布-

平成 28 年 9 月 11 日

概 要

PRMLの「10.1.3 例:一変数ガウス分布」についての実装と考察

目 次

1　問題設定

ガウス分布に対して, 分解による変分近似を行う.

2　アルゴリズム

同一のガウス分布から独立に発生したと仮定するデータ集合 $D = \{x_1, \ldots, x_n\}$ が与えられたとき, 元のガウス分布の平均 μ と精度 τ の事後分布を求めることが目標である.
まず, 尤度関数は

$$p(D|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{N/2} \exp\left\{-\frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\} \quad (10.21)$$

となり, パラメータの事前分布は

$$p(\mu|\tau) = N(\mu|\mu_0, (\lambda_0\tau)^{-1}) \quad (10.22)$$

$$p(\tau) = Gam(\tau|a_0, b_0) \quad (10.23)$$

と, ガウス-ガンマ共役事前分布が得られ, 事後分布は厳密な形で求めることができ, これもガウス-ガンマ分布になる. ただし, 今はこれを使わず, 事後分布を分解した変分近似を考える.

$$q(\mu,\tau) = q_\mu(\mu)q_\tau(\tau) \quad (10.24)$$

となるが, この各因子をそれぞれ推測したい. これには次の

$$\ln q_j^*(\mathbf{z}_j) = E_{i\neq j}[\ln p(X,Z)] + const \quad (10.9)$$

を用いる.
まず, $\ln q_\mu(\mu)$ については

$$\begin{aligned}\ln q_\mu^*(\mu) &= E_\tau[\ln p(D|\mu,\tau) + \ln p(\mu|\tau)] + const \\ &= -\frac{E[\tau]}{2}\left\{\lambda_0(\mu-\mu_0)^2 + \sum_{n=1}^{N}(x_n-\mu)^2\right\} + const \quad (10.25)\end{aligned}$$

で, これは平方完成することにより $q_\mu(\mu)$ がガウス分布になることを示す.

$$q_\mu(\mu) = N(\mu|\mu_N, \lambda_N^{-1})$$

ただし,

$$\mu_N = \frac{\lambda_0\mu_0 + N\bar{x}}{\lambda_0 + N} \quad (10.26), \quad \lambda_N = (\lambda_0 + N)E[\tau] \quad (10.27)$$

となる.
また, 同様にして $\ln q_\tau(\tau)$ については

$$\begin{aligned}\ln q_\tau^*(\tau) &= E_\mu[\ln p(D|\mu,\tau) + \ln p(\mu|\tau)] + const \\ &= (a_0-1)\ln\tau - b_0\tau + \frac{N+1}{2}\ln\tau - \frac{\tau}{2}E_\mu[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2] + const \quad (10.28)\end{aligned}$$

で, これは $q_\tau(\tau)$ がガンマ分布になることを示す.

$$q_\tau(\tau) = Gam(\tau|a_N, b_N)$$

ただし,

$$a_N = a_0 + \frac{N+1}{2} \quad (10.29), \quad b_N = b_0 + \frac{1}{2} E_\mu [\sum_{n=1}^{N} (x_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2] \quad (10.30)$$

となる.
そして, これらを計算するためには, $\ln q_\mu^*(\mu)$ と $\ln q_\tau^*(\tau)$ のパラメータが相互に関係していることから, 繰り返しを必要とする.

変分推論

1. ガウス分布では, 平均 μ, 精度 τ の共役事前分布にガウス分布, ガンマ分布を選ぶ.
2. 必要となるモーメント $E[\tau]$ を適当に初期化する.
3. 因子 $q_\mu^*(\mu) = N(\mu|\mu_N, \lambda_N)$ のパラメータを推定する.
4. 因子 $q_\tau^*(\tau) = Gam(\tau|a_N, b_N)$ のパラメータを推定する.
5. 3,4 を繰り返す.

3 コード

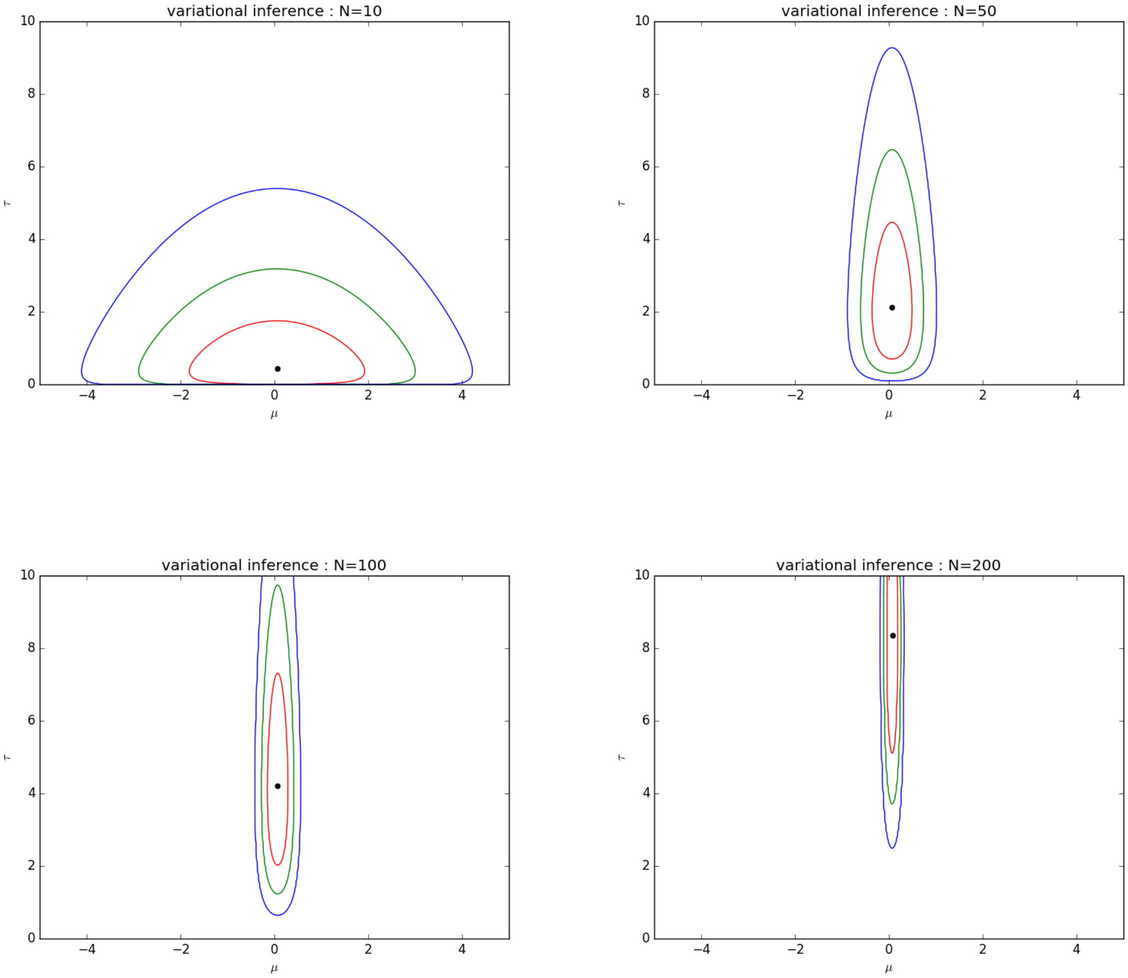
ガウス分布に対する変分近似のコード (variation_gauss.py).

```
def gauss(x,mu,lam):
  return sqrt(lam/2/pi)*exp(-(x-mu)**2*lam/2)

def gamma(x,a,b):
  return (b**a)*np.power(x,a-1)*exp(-b*x)/gam(a)

"""データセットの作成"""
dataset="dataset.csv"
data=sp.genfromtxt(dataset,delimiter=",")

for N in [10,50,100,200]:
  x=data
  """変分近似"""
  #超パラメータ(a0, b0の決め方がわからない)
  mu0,lam0,a0,b0=0,3,0,0
  #モーメント
  Ex=mean(x)
  Ex2=mean(x*x)
  Etau=mean(x*x)-mean(x)**2
  #変分推論
  diff=100
  muN=(lam0*mu0+N*Ex)/(lam0+N)
  aN=a0+(N+1)/2
  lamN,bN=lam0,b0
  while diff>=10**-3:
    tmp1,tmp2=lamN,bN
    lamN=(lam0+N)*Etau
    bN=b0+(Ex2 - 2*(Ex+lam0*mu0)*muN + (1+lam0)*(1/lamN+Ex**2) + lam0*mu0**2)/2
    Etau=aN/bN
    diff=abs(tmp1-lamN)+abs(tmp2-bN)
```

4 結果

平均 0, 分散 25 のガウス分布に従うデータセット ($N = 10, 50, 100, 200$) に対して, 変分近似を行った.

図 1: $N = 10, 50, 100, 200$



5 まとめ

N が大きくなるにつれて, 平均, 分散 (精度の逆数) は絞り込めていることがわかる.
まだ, 超パラメータについても考える必要がある.