変分ベイズ推定

B8EM1016　富田優

　変分ベイズ推定(Variational Bayesian(VB) method)とは、確率モデルの潜在変数や事後分布のパラメータを求める手法。確率モデルを直接解くのではなく、因子分解という仮定を置くことで計算を容易にしている。積分の近似としてサンプリングを用いる MCMC と異なり、パラメータを解析的に解いているので決定論的である(データを与えれば一意にパラメータが求まる)。
以上の 2 点(①因子分解という過程　②解析的に解く)をもって、決定論的な近似アルゴリズムと言われる。

以下では変分ベイズ推定を説明する準備として、
　1.KL 情報量
　2.変分法
を説明する。

1.KL 情報量

統計モデルの「近さ」を表す指標としてカルバック・ライブラー情報量がある。

$$KL[p^*(x)||p(x|\Phi)] = \int p^*(x)\log\frac{p^*(x)}{p(x|\Phi)}dx$$

KL 情報量には以下の性質がある。

$$KL[p^*(x)||p(x|\Phi)] \geq 0$$
$$KL[p^*(x)||p(x|\Phi)] = 0 \leftrightarrow p^*(x) = p(x|\Phi)$$

2.変分法

変分法とは、関数fを入力とする汎関数$\mathrm{L}[f(x)]$の極値となる関数fを求めるための方法である。
つまり、

$$\frac{\partial \mathrm{L}[f]}{\partial f} = 0$$

となる関数を求める方法である。
ここで、関数$\mathrm{f}(x)$が関数$\mathrm{q}(x)$によって構成されている場合に、明示的に$\mathrm{f}(x,q)$と書くことにする。例えば、$\mathrm{f}(x) = -q(x)\log q(x)$という関数を$\mathrm{f}(x,q)$などと書く。この表現を用いて汎関数を以下のように定義する。

$$L[q(x)] = \int f(x,q)\,dx$$

この汎関数の極値を与えるは、以下のオイラー・ラグランジュ方程式を解けばよい。

$$\frac{\partial f(x,q)}{\partial q(x)} = 0$$

3.変分ベイズ推定

一般的なモデルにおける変分ベイズ法の導出を説明する。
モデルの生成過程

$$x_i \sim \mathrm{p}\left(x_i \middle| \varphi_{z_i}\right), z_i \sim Multi(z_i|\pi), \varphi_k \sim p(\varphi_k|\eta), \pi \sim Dir(\pi|\alpha).$$

3.1 事後分布の計算

求めたい事後分布$\mathrm{q}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi})$を解析的に求めることは困難なので、**因子分解**が可能という過程を置いて、以下の近似事後分布を解くことを考える。

$$\mathrm{q}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi}) = \prod_{i=1}^{n} q(z_i) \prod_{k=1}^{K} q(\varphi_k)\, q(\boldsymbol{\pi})$$

また、確率分布間の近さとして KL 情報量を用いれば、近似事後分布の推定は以下の最適化問題として定式化できる。

$$q^*(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi}) = \underset{\mathrm{q}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi}) \in \Omega}{\operatorname{argmin}} KL[\mathrm{q}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi}) || \mathrm{p}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi} | \boldsymbol{x_{1:n}}, \boldsymbol{\alpha}, \boldsymbol{\eta})]$$

しかし、上記の式には計算の困難な$\mathrm{p}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi} | \boldsymbol{x_{1:n}}, \boldsymbol{\alpha}, \boldsymbol{\eta})$が含まれているため、効率的に解くことができない。この問題を直接最適化することは難しいが、以下の対数周辺尤度を利用して、最適化問題の式を直接解かずに$\mathrm{q}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi})$を導出できる。

$$\log p(x_{1:n}|\alpha,\eta) = \log \int \sum_{z_{1:n}} p(x_{1:n}, z_{1:n}, \varphi, \pi | \alpha, \eta) d\varphi d\pi$$

3.2 対数周辺尤度の展開

$$\log p(z_{1:n}|\alpha,\eta) = \mathrm{F}[q(z_{1:n}, \varphi, \pi)] + KL[\mathrm{q}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi}) || \mathrm{p}(\boldsymbol{z_{1:n}}, \boldsymbol{\varphi}, \boldsymbol{\pi} | \boldsymbol{x_{1:n}}, \boldsymbol{\alpha}, \boldsymbol{\eta})]$$

ただし、

$$\mathrm{F}[q(z_{1:n}, \varphi, \pi)] = \int \sum_{z_{1:n}} q(z_{1:n}, \varphi, \pi) \log \frac{p(x_{1:n}, z_{1:n}, \varphi, \pi | \alpha, \eta)}{q(z_{1:n}, \varphi, \pi)} d\varphi d\pi$$

である。

3.3 イエンセンの不等式

ここで、イエンセンの不等式

$$\int f\big(y(x)\big)\,p(x)\,\mathrm{dx} \geq f\left(\int y(x)p(x)\,dx\right)$$

$\mathrm{p}(x)$：実数上の確率分布　$\mathrm{y}(x)$：実数上の可積分関数　$\mathrm{f}(x)$：実数上の凸関数

を用いると

$$\log p(\boldsymbol{x_{1:n}}|\alpha,\eta) \geq F[q(\boldsymbol{z_{1:n}},\boldsymbol{\varphi},\boldsymbol{\pi})]$$

である。

以上の式のとおり、対数周辺尤度の下限という意味で$\mathrm{F}[q(z_{1:n},\varphi,\pi)]$は変分下限(Variational lower-bound)と呼ばれている。

ここで重要なことは

$\log p(z_{1:n}|\alpha,\eta)$が$q(z_{1:n},\varphi,\pi)$とは無関係なことである。変分下限$\mathrm{F}[q(z_{1:n},\varphi,\pi)]$を最大にする$q(z_{1:n},\varphi,\pi)$は KL 情報量を最小にする$q(z_{1:n},\varphi,\pi)$である。

3.4 近似事後分布が帰着する最適化問題

最終的に、変分ベイズ法は以下の最適化問題に帰着する。

$$q^*(\boldsymbol{z_{1:n}},\boldsymbol{\varphi},\boldsymbol{\pi}) = \underset{\mathrm{q}(\boldsymbol{z_{1:n}},\boldsymbol{\varphi},\boldsymbol{\pi})\in\Omega}{\operatorname{argmax}} \mathrm{F}[q(z_{1:n},\varphi,\pi)]$$

さらに細かく$\mathrm{F}[q(z_{1:n},\varphi,\pi)]$の展開を行う。

$$\mathrm{F}[q(z_{1:n},\varphi,\pi)] = \int\sum_{z_{1:n}} q(z_{1:n})q(\varphi)q(\pi)\log\frac{p(x_{1:n},z_{1:n}|\varphi,\pi)p(\varphi|\eta)p(\pi|\alpha)}{q(z_{1:n})q(\varphi)q(\pi)}d\varphi d\pi$$

$$= \int\sum_{z_{1:n}} q(z_{1:n})q(\varphi)q(\pi)\log\frac{p(x_{1:n},z_{1:n}|\varphi,\pi)}{q(z_{1:n})}d\varphi d\pi + \sum_{k=1}^{K}\int q(\varphi_k)\log\frac{q(\varphi_k|\eta)}{q(\varphi_k)}d\varphi_k$$

$$+ \int q(\pi)\log\frac{q(\pi|\alpha)}{q(\pi)}d\pi$$

$$= \int\sum_{z_{1:n}} q(z_{1:n})q(\varphi)q(\pi)\log\frac{p(x_{1:n},z_{1:n}|\varphi,\pi)}{q(z_{1:n})}d\varphi d\pi - \sum_{k=1}^{K}KL\big[q\big(\varphi_k||p(\varphi_k|\eta)\big)\big]$$

$$- KL[q(\pi)||p(\pi|\alpha)]$$

これで、求めたい$q(z_{1:n}),q(\varphi),q(\pi)$を求めやすい形にすることができた。

3.5 変分法で解く

それでは、上記の式に変分法を用いて解く。

まずは、右辺第一項を展開すると

$$\int \sum_{z_{1:n}} q(z_{1:n}) q(\varphi) q(\pi) \log \frac{p(x_{1:n}, z_{1:n} | \varphi, \pi)}{q(z_{1:n})} d\varphi d\pi$$

$$= \int \sum_{z_{1:n}} q(z_{1:n}) q(\varphi) q(\pi) \log p(x_{1:n}, z_{1:n} | \varphi, \pi)\, d\varphi d\pi - \int q(\varphi) q(\pi) d\varphi d\pi \sum_{z_{1:n}} q(z_{1:n}) \log q(z_{1:n})$$

$$= \int \sum_{i=1}^{n} \sum_{z_i} q(z_i) q(\varphi) q(\pi) \log p(x_i, z_i | \varphi, \pi)\, d\varphi d\pi - \sum_{i=1}^{n} \sum_{k=1}^{K} q(z_i = k) \log p(z_i = k)$$

となる。

ここで、$\mathrm{F}[q(z_{1:n}, \varphi, \pi)]$の$\mathrm{q}(z_i)$に関係のある項のみを取り出して、それを$\tilde{F}[q(z_i)]$とおくと

$$\tilde{F}[q(z_i)] = \sum_{k=1}^{K} q(z_i = k) \int q(\varphi)\, q(\pi) \log p(x_i, z_i = k | \varphi, \pi) d\varphi d\pi - \sum_{k=1}^{K} q(z_i = k) \log p(z_i = k)$$

となる。$q(z_i = k)$を変数とみなすと、上記の式は通常の微分法で解ける。また、

$$\sum_{k=1}^{K} q(z_i = k) = 1$$

となることを考慮すると、これは制約付き最適化問題なので、ラグランジュ未定乗数法を用いてこれを解くことになる。

式

$$\mathrm{L}[q(z_i)] = \tilde{F}[q(z_i)] + \lambda \left(1 - \sum_{k=1}^{K} q(z_i = k) \right)$$

を$q(z_i)$で微分すると、

$$\int q(\varphi)\, q(\pi) \log p(x_i, z_i = k | \varphi, \pi) d\varphi d\pi - \log p(z_i = k) - q(z_i = k) \frac{1}{q(z_i = k)} - \lambda = 0$$

$$\leftrightarrow q(z_i = k) = exp\left[-1 - \lambda + \int q(\varphi)\, q(\pi) \log p(x_i, z_i = k | \varphi, \pi) d\varphi d\pi \right]$$

ラグランジュ関数をλで微分して得られた

$$\sum_{k=1}^{K} q(z_i = k) = 1$$

に上の式を代入すると

$$q(z_i = k) = \frac{exp[\int q(\varphi)\, q(\pi) \log p(x_i, z_i = k | \varphi, \pi) d\varphi d\pi]}{\sum_{k'=1}^{K} exp[\int q(\varphi)\, q(\pi) \log p(x_i, z_i = k' | \varphi, \pi) d\varphi d\pi]}$$

となる。

○**ポイント**

「事後分布の計算を結合分布の計算に帰着させる」

変分ベイズ法では、変分下限という形で、事後分布の計算を結合分布の計算に帰着させている

4.LDA の変分ベイズ法

要点

目的：KL$[p(\boldsymbol{z},\boldsymbol{\varphi},\boldsymbol{\theta}|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})||q(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\varphi})]$を最小にする$q(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\varphi})$を求める

手法：対数周辺尤度$p(\boldsymbol{w}|\boldsymbol{\alpha},\boldsymbol{\beta})$の変分下限$\mathrm{F}[\boldsymbol{q}(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\varphi})]$を導出し、$\mathrm{F}[\boldsymbol{q}(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\varphi})]$を最大にする$q(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\varphi})$を変分法により求める。

変分下限の導出プロセスは

1.周辺化された確率変数を結合分布として積分系で表示する。

2.変分事後分布を分子分母でキャンセルする形で導入する。

3.イエンセンの不等式により下限を求める。

$$\log p(w|\alpha,\beta) = \log \int \sum_z p(w,z,\varphi,\theta|\alpha,\beta)d\varphi d\theta$$

$= \log \int \sum_z q(z,\theta,\varphi)\frac{p(w,z,\varphi,\theta|\alpha,\beta)}{q(z,\theta,\varphi)}d\varphi d\theta$

$$\geq \int \sum_z q(z,\theta,\varphi)\log\frac{p(w,z,\varphi,\theta|\alpha,\beta)}{q(z,\theta,\varphi)}d\varphi d\theta \equiv \mathrm{F}[q(z,\theta,\varphi)]$$

ここで近似事後分布として

$$q(z,\theta,\varphi) = \left[\prod_{d=1}^{M}\prod_{i=1}^{n_d} q(z_{d,i})\right]\left[\prod_{d=1}^{M} q(\theta_d)\right]\left[\prod_{k=1}^{K} q(\varphi_k)\right]$$

以上の因子分解の仮定を置く。

$$\mathrm{F}[q(z,\theta,\varphi)] = \int \sum_z q(z)\,q(\theta)\,q(\varphi)\log p(w|z,\varphi)\,p(z|\theta)d\theta d\varphi - \sum_z q(z)\log q(z)$$

$$-\int q(\theta)\log\frac{p(\theta|\alpha)}{q(\theta)}d\theta - \int q(\varphi)\log\frac{p(\varphi|\beta)}{q(\varphi)}d\varphi$$

$$= \int \sum_{d=1}^{M} \sum_{i=1}^{n_d} q(z_{d,i})\, q(\theta_d)\, q(\varphi) \log p(w_{d,i} | z_{d,i}, \varphi)\, p(z_{d,i} | \theta_d) d\varphi d\theta$$

$$- \sum_{d=1}^{M} \sum_{i=1}^{n_d} \sum_{k=1}^{K} q(z_{d,i} = k) \log q(z_{d,i} = k) + \sum_{d=1}^{M} - \int q(\theta_d) \log \frac{p(\theta_d | \alpha)}{q(\theta_d)} d\theta_d$$

$$+ \sum_{k=1}^{K} - \int q(\varphi_k) \log \frac{p(\varphi_k | \beta)}{q(\varphi_k)} d\varphi_k$$

以下、$\mathrm{F}[q(z,\theta,\varphi)]$を最大にする$q(\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\varphi})$を変分法に求める。

まず、$\mathrm{F}[q(z,\theta,\varphi)]$から $q(\theta_d)$に関係する項のみを抜き出した$\tilde{F}[q(\theta_d)]$は

$$\tilde{F}[q(\theta_d)] = \int q(\theta_d) \sum_{z} q(z) \sum_{i=1}^{n_d} \log p(z_{d,i} | \theta_d)\, d\theta_d - \int q(\theta_d) \log \frac{q(\theta_d)}{p(\theta | \alpha)} d\theta_d$$

である。

汎関数$\tilde{F}[q(\theta_d)] = \int f(\theta_d, q(\theta_d))\, d\theta_d$と見れば、

$$\frac{\delta \tilde{F}[q(\theta_d)]}{\delta q(\theta_d)} = 0$$

となる$q(\theta_d)$は、変分法により

$$\frac{\partial f(\theta_d, q(\theta_d))}{\partial q(\theta_d)} = 0$$

となる$q(\theta_d)$をお願い致します。求めればよいので

$$\frac{\partial f(\theta_d, q(\theta_d))}{\partial q(\theta_d)} = \sum_{z} q(z) \sum_{i=1}^{n_d} \log p(z_{d,i} | \theta_d) - \log \frac{q(\theta_d)}{p(\theta_d | \alpha)} - 1 = 0$$

$$\leftrightarrow q(\theta_d) \propto p(\theta_d | \alpha) exp \left[\sum_{z} q(z) \sum_{i=1}^{n_d} \log p(z_{d,i} | \theta_d) \right]$$

ここで、

$$q(\theta_d | \alpha) \propto \prod_{k=1}^{K} \theta_{d,k}^{\alpha_k - 1}$$

$$E_{q(z_d)}[n_{d,k}] = \sum_{i=1}^{n_d} q(z_{d,i} = k)$$

$$\mathrm{p}(z_{d,i} | \theta_d) = \prod_{k=1}^{K} \theta_{d,k}^{\delta(z_{d,i} = k)}$$

を用いると

$$\propto \prod_{k=1}^{K} \theta_{d,k}^{\alpha_k - 1} exp \left[\sum_{z} q(z) \sum_{i=1}^{n_d} \sum_{k=1}^{K} \delta(z_{d,i} = k) \log \theta_{d,k} \right]$$

$$= \exp\left[\sum_{k=1}^{K}(\alpha_k - 1)\log\theta_{d,k}\right] exp\left[\sum_{k=1}^{K}\sum_{i=1}^{n_d} q\left(z_{d,i} = k\right)\log\theta_{d,k}\right]$$

$$= \exp\left[\sum_{k=1}^{K}(\alpha_k - 1)\log\theta_{d,k}\right] exp\left[\sum_{k=1}^{K} E_{q(z_d)}\left[n_{d,k}\right]\log\theta_{d,k}\right]$$

$$= \exp\left[\sum_{k=1}^{K}\left(E_{q(z_d)}\left[n_{d,k}\right] + \alpha_k - 1\right)\log\theta_{d,k}\right]$$

$$= \prod_{k=1}^{K}\theta_{d,k}^{E_{q(z_d)}\left[n_{d,k}\right]+\alpha_k-1}$$

ここで

$$\xi_{d,k}^{\theta} = E_{q(z_d)}\left[n_{d,k}\right] + \alpha_k$$

とおくと式から、$q(\theta_d)$は$\xi_d^{\theta} = \left(\xi_{d,1}^{\theta}, \xi_{d,2}^{\theta}, \cdots, \xi_{d,k}^{\theta}\right)$をパラメータとするディリクレ分布であることが分かる。従って、正規化項の計算も容易で、

$$q\left(\theta_d | \xi_d^{\theta}\right) = \frac{\Gamma\left(\sum_{k=1}^{K}\xi_{d,k}^{\theta}\right)}{\prod_{k=1}^{K}\Gamma\left(\xi_{d,k}^{\theta}\right)}\prod_{k=1}^{K}\theta_{d,k}^{\xi_{d,k}^{\theta}-1}$$

と求めることができる。

引用：トピックモデルによる統計的潜在意味解析：奥村学、佐藤一誠(2013)