10.3 変分線形回帰

平成 28 年 9 月 11 日

概 要

PRML の「10.3 変分線形回帰」についての実装と考察

目 次

1 問題設定

線形回帰を行う．超パラメータ α, β についてエビデンス近似ではなく変分近似で近似してみる．

2 アルゴリズム

線形回帰では, 超パラメータまで最適化しようとすると, 解析的に解くことのできない積分が出てきた．今まで, これをエビデンス近似で対処してきたが, ここでは変分近似を用いてみる.(ただし, β は既知とする)
まず, $\mathbf{w}$ に対する尤度関数と W の事前分布は,

$$p(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N} N(t_n|\mathbf{w}^T\boldsymbol{\phi}_n, \beta^{-1}) \quad (10.87)$$

$$p(\mathbf{w}|\alpha) = N(\mathbf{w}|\mathbf{0}, \alpha^{-1}I) \quad (10.88)$$

とする．また, α の事前分布は共役事前分布として

$$p(\alpha) = G(\alpha|a_0, b_0) \quad (10.89)$$

を選ぶ．以上より, すべての変数の同時分布は

$$p(\mathbf{t}, \mathbf{w}, \alpha) = p(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\alpha)p(\alpha) \quad (10.90)$$

となる．
ここでの目標は事後分布 $p(\mathbf{w}, \alpha|\mathbf{t})$ の近似を求めることなので, このため分解される変分事後分布を考える．

$$q(\mathbf{w}, \alpha) = q(\mathbf{w})q(\alpha) \quad (10.91)$$

まず, $\ln q(\alpha)$ は

$$\begin{aligned}
\ln q^*(\alpha) &= \ln p(\alpha) + E_{\mathbf{w}}[\ln p(\mathbf{w}|\alpha)] + const \\
&= (a_0 - 1)\ln\alpha - b_0\alpha + \frac{M}{2}\ln\alpha - \frac{\alpha}{2}E[\mathbf{w}^T\mathbf{w}] + const \quad (10.92)
\end{aligned}$$

これは, ガンマ分布の対数であり

$$q^*(\alpha) = Gam(\alpha|a_N, b_N) \quad (10.93)$$

となる．ここで

$$a_N = a_0 + \frac{M}{2} \quad (10.94), \quad b_N = b_0 + \frac{1}{2}E[\mathbf{w}^T\mathbf{w}] \quad (10.95)$$

とした．
また, $\ln q(\mathbf{w})$ は

$$\begin{aligned}
\ln q^*(\mathbf{w}) &= \ln p(\mathbf{t}|\mathbf{w}) + E_{\alpha}[\ln p(\mathbf{w}|\alpha)] + const \\
&= -\frac{\beta}{2}\sum_{n=1}^{N}\{\mathbf{w}^T\boldsymbol{\phi}_n - t_n\}^2 - \frac{1}{2}E[\alpha]\mathbf{w}^T\mathbf{w} + const \\
&= -\frac{1}{2}\mathbf{w}^T(E[\alpha]I + \beta\Phi^T\Phi)\mathbf{w} + \beta\mathbf{w}^T\Phi^T\boldsymbol{t} + const
\end{aligned}$$

これは, ガウス分布の対数であり

$$q^*(\mathbf{w}) = N(\mathbf{w}|\mathbf{m}_N, S_N) \quad (10.99)$$

となる. ここで

$$\mathbf{m}_N = \beta S_N \Phi^T \mathbf{t} \quad (10.100), \quad S_N = (E[\alpha]I + \beta\Phi^T\Phi)^{-1} \quad (10.101)$$

とした.
これらの計算には

$$E[\alpha] = \frac{a_N}{b_N} \quad (10.102), \quad E[\mathbf{w}^T\mathbf{w}] = \mathbf{m}_N\mathbf{m}_N^T + S_N \quad (10.103)$$

を用いて, 交互に再推定していく.

変分線形回帰

1. まず $q^*(\alpha)$ のパラメータを初期化する.
2. $q^*(\mathbf{w})$ のパラメータを推定する.
3. $q^*(\alpha)$ のパラメータを推定する.
4. 2,3 を繰り返す.

3 コード

変分線形回帰のコード (variational_linear_regression.py).

```
for N in [20,100,500]:
        x=data[:N,0]
        t=data[:N,1]
        for M in [4,10,20]:
                mu=[(m+0.5)*(2*pi/(M+1)) for m in range(M)]
                s=(2*pi)**2/12
                P=np.zeros((N,M))
                for n in range(N):
                        for m in range(M):
                                P[n,m]=polynomial_basis(x[n],m)
                                #tanh_basis(x[n],m,mu[m],s)
                I=np.identity(M)

                beta=1/(0.3)**2
                a0,b0=1,1
                aN,bN=a0+M/2,1
                mN=np.zeros(M)
                SN=np.zeros((M,M))

                diff=10**6
                while diff>=10**-6:
                        diff=0
                        tmp=SN
                        SN=inv(aN/bN*I+beta*dot(P.T,P))
                        diff+=norm(tmp-SN)
                        tmp=mN
                        mN=beta*np.dot(SN,dot(P.T,t))
                        diff+=norm(tmp-mN)
                        tmp=bN
                        bN=b0+trace(outer(mN,mN)+SN)/2
                        diff+=abs(tmp-bN)
```

4　結果

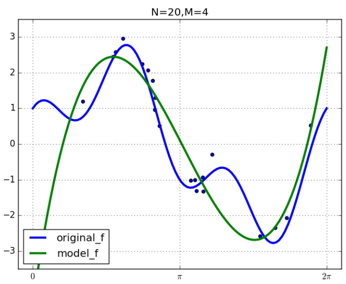
$\alpha = 10^{-10}$,　$\beta = 1.0/(0.3)^2$ に固定して実験を行った.
この β は元のデータの分散が $(0.3)^2$ であることに由来する. α については適当.

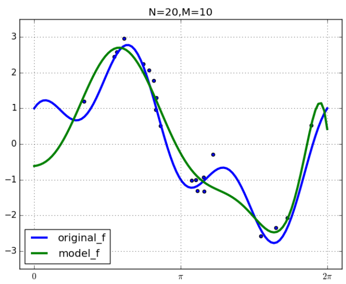
4.1　多項式基底

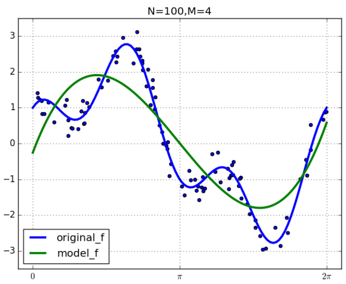
基底 $\Phi(x)$ に $\phi_i(x) = x^i$ を選んだ. 「多項式曲線フィッティング」で行ったものと等しい.

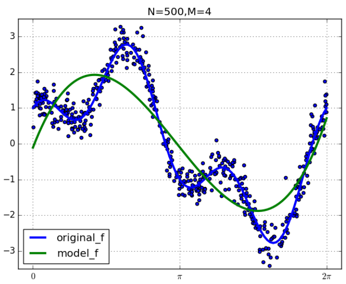
N M	20	100	500
4	1.08	0.81	0.80
10	0.67	-	-
20	-	-	-

表 1: E_{RMS} の N,M との関係 (多項式基底)









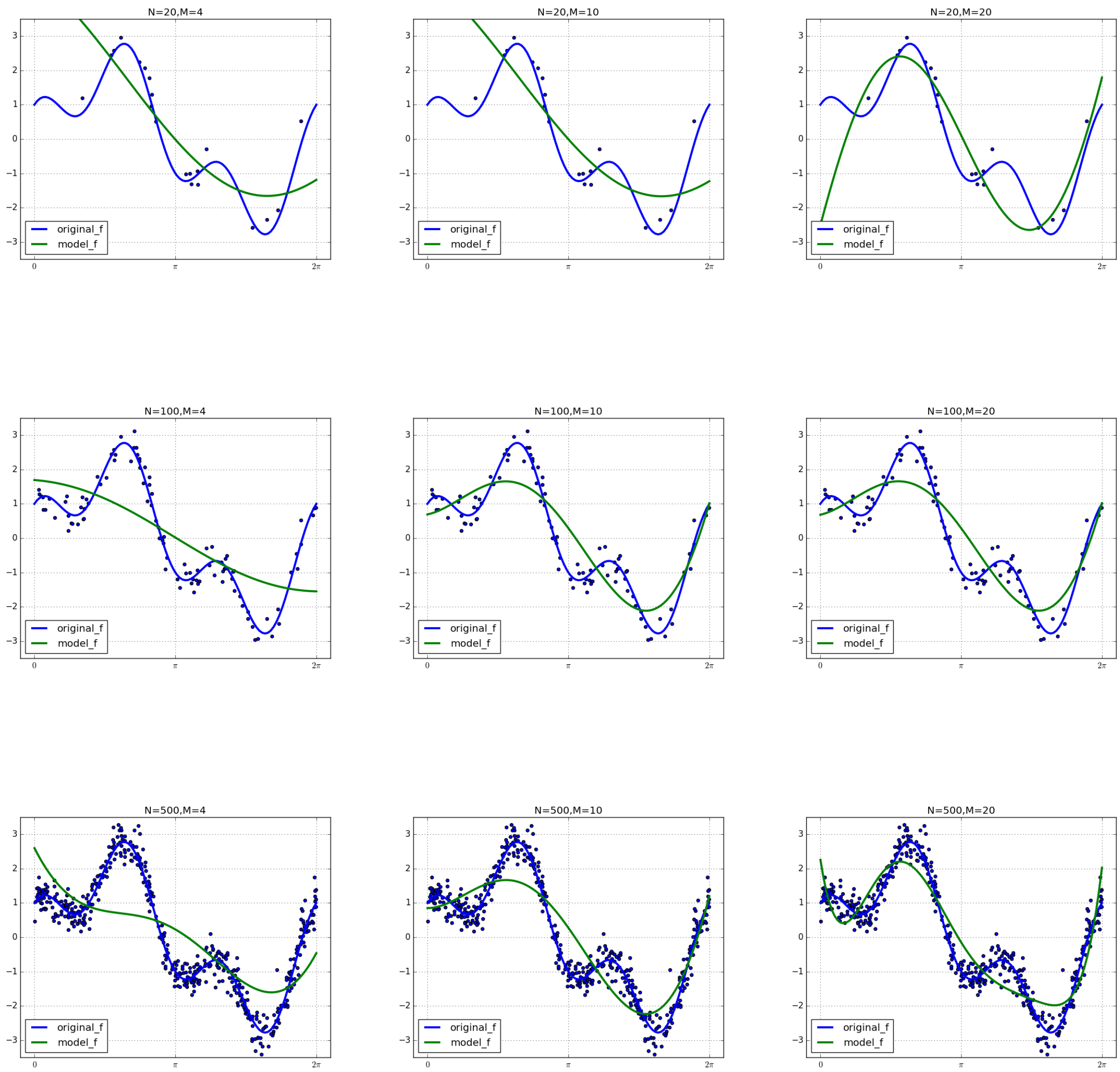
4.2　ガウス基底

基底 $\Phi(x)$ に $\phi_i(x) = exp\{-\frac{(x-\mu_i)^2}{2s^2}\}$ をもちいる．またここでは μ は区間 $[0, 2\pi]$ を M 個に等分する区間の中心を用い,s には x の分散を用いた.

$$mu = [(m + 0.5) * (2 * pi/(M + 1))\ \ for\ m\ in\ range(M)]\quad s = (2 * pi)/M$$

M \ N	20	100	500
4	1.68	1.07	1.00
10	1.68	0.74	0.74
20	0.94	0.74	0.62

表 2: E_{RMS} の N,M との関係 (ガウス基底)

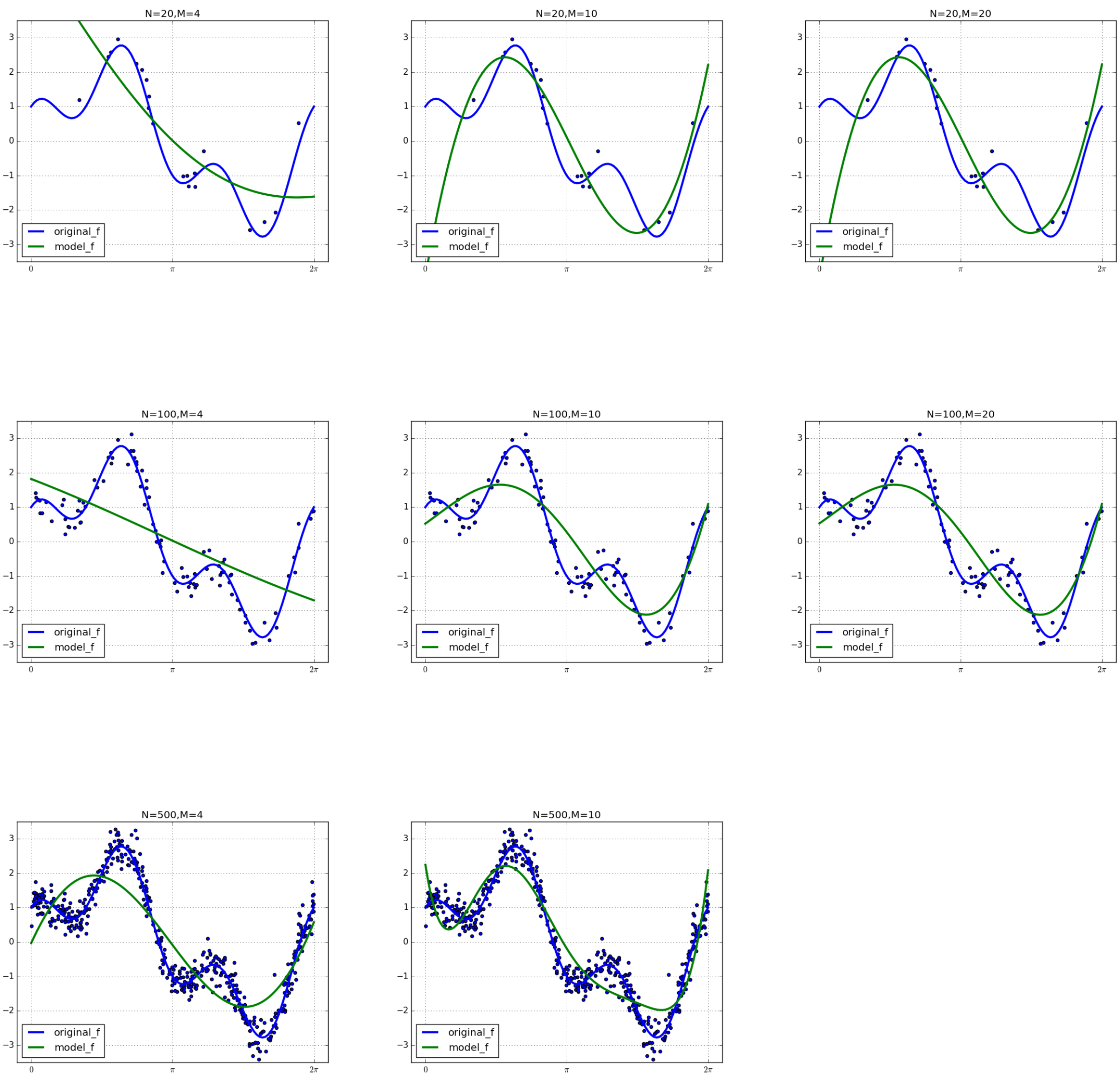


4.3 シグモイド基底

基底 $\Phi(x)$ に $\phi_i(x) = \sigma(\frac{x-\mu_i}{s})$ をもちいる. ただし,$\sigma(a) = \frac{1}{1+e^{-a}}$

M \ N	20	100	500
4	1.82	1.14	0.80
10	1.05	0.74	0.61
20	1.05	0.74	-

表 3: E_{RMS} の N,M との関係 (シグモイド基底)

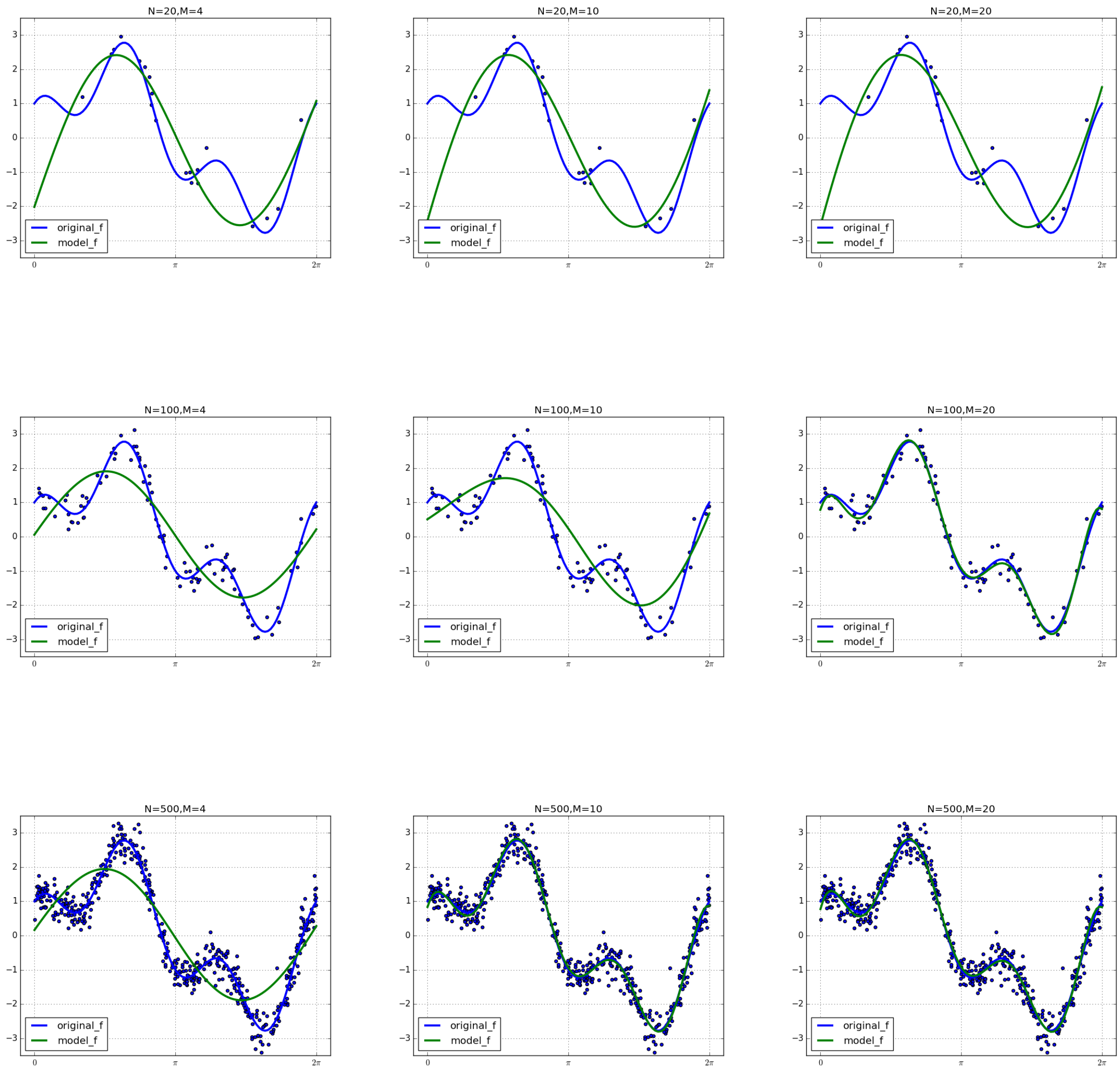


4.4 tanh 基底

基底 $\Phi(x)$ に $\phi_i(x) = tanh(\frac{x-\mu_i}{s})$ をもちいる.

M \ N	20	100	500
4	0.92	0.81	0.81
10	0.94	0.77	0.30
20	0.95	0.31	0.30

表 4: E_{RMS} の N,M との関係 (tanh 基底)



5 まとめ

毎回のように, $tanh$ 関数を基底関数とするのが一番よく, 多項式基底はあまり良くない. 繰り返しを要するので収束しないこともあった.