6.4.1〜6.4.2 ガウス過程による回帰

平成 28 年 9 月 11 日

概 要

PRMLの「6.4.1 線形回帰再訪」「6.4.2 ガウス過程による回帰」についての実装と考察.

目 次

1 問題設定

t_n がガウス分布に従っているものをガウス過程という. このガウス過程について考察する.
ここでは, ガウス過程による回帰を行う.

2 アルゴリズム

まず, 観測される目標変数の値にはノイズが含まれるとする.

$$t_n = y_n + \epsilon_n \quad (6.57)$$

そして, このノイズはガウス分布に従うとする.

$$p(t_n|y_n) = N(t_n|y_n, \beta^{-1}) \quad (6.58)$$

これを N 個の観測地についてまとめると

$$p(\mathbf{t}|\mathbf{y}) = N(\mathbf{t}|\mathbf{y}, \beta^{-1}I) \quad (6.59)$$

となる.
また, ガウス過程の過程から $p(\boldsymbol{y})$ は

$$p(\mathbf{y}) = N(\mathbf{y}|\mathbf{0}, K) \quad (6.60)$$

となる. ここで, K はグラム行列. ただし, (i,j) 成分は x_i, x_j の相関度合いによって決まるものである必要がある.
入力 $x_1, \ldots, x_N$ で条件づけられた周辺確率分布 $p(\mathbf{t})$ は (線形ガウスモデル)

$$p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{y})p(\mathbf{y})d\mathbf{y} = N(\mathbf{t}|\mathbf{0}, C) \quad (6.61)$$

ここで共分散行列 C は

$$C(x_n, x_m) = k(x_n, x_m) + \beta^{-1}\delta_{nm} \quad (6.62)$$

また, カーネル関数には

$$k(x_n, x_m) = \theta_0 exp\{-\frac{\theta_1}{2}\|x_n - x_m\|^2\} + \theta_2 + \theta_3 x_n^T x_m \quad (6.63)$$

がよく用いられる.
次に, 条件付き分布 $p(t_{N+1}|\mathbf{t})$ を求めるが, まず $p(\mathbf{t}_{N+1})$ については (6.61) と同様にして

$$p(\mathbf{t}_{N+1}) = N(\mathbf{t}_{N+1}|\mathbf{0}, C_{N+1}) \quad (6.64)$$

となり, 共分散行列は

$$C_{N+1} = \begin{pmatrix} C_N & \boldsymbol{k} \\ \boldsymbol{k}^T & c \end{pmatrix} \quad (6.65)$$

で, $\boldsymbol{k}$ は成分 n に $k(x_n, x_{N+1})$ に持つベクトルで, スカラー $c = k(x_{N+1}, x_{N+1}) + \beta^{-1}$ とする.
分割されたガウス分布についての結果から

$$m(x_{N+1}) = \mathbf{k}^T C_N^{-1}\boldsymbol{t} \quad (6.66)$$

$$\sigma^2(x_{N+1}) = c - \mathbf{k}^T C_N^{-1}\mathbf{k} \quad (6.67)$$

これは x_{N+1} に対する予測分布のパラメータととらえられる.

3 コード

予測分布のパラメータを求める (kaiki.py).

```
"""カーネル関数の定義"""
theta=0.05
def gauss(x,z):
        return np.exp(-(x-z)**2/2*N*theta)

"""Wの最適化"""
print("gauss")
for N in [10,30,50,100]:
        x=data[:N,0]
        t=data[:N,1]
        C=np.identity(N)*(0.3**2)
        s=2*(2*pi)/N

        for n in range(N):
                for m in range(N):
                        C[n,m]+=gauss(x[n],x[m])

        #x=piに対する予測分布の平均
        print(model_f(pi))
        #x=piに対する予測分布の分散
        k=[gauss(x[n],pi) for n in range(N)]
        print(gauss(pi,pi)+1/0.3**2-dot(k,dot(inv(C),k)))

        #求まったパラメータからモデル関数を作り
        def model_f(z):
                k=[gauss(x[n],z) for n in range(N)]
                return dot(k,dot(inv(C),t))
```

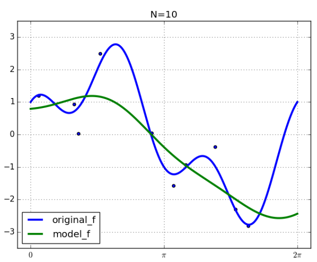
4 結果

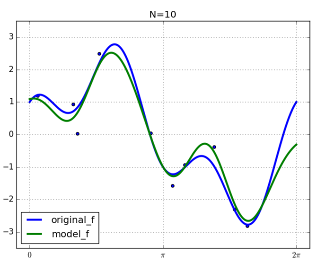
4.1 ガウスカーネル

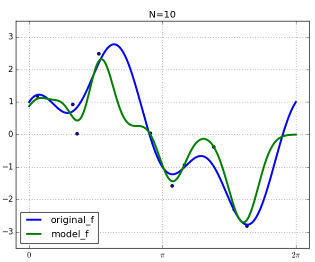
予測分布の平均をプロットする.
ここでは, $\beta = 1/0.3^2$ としカーネル関数に

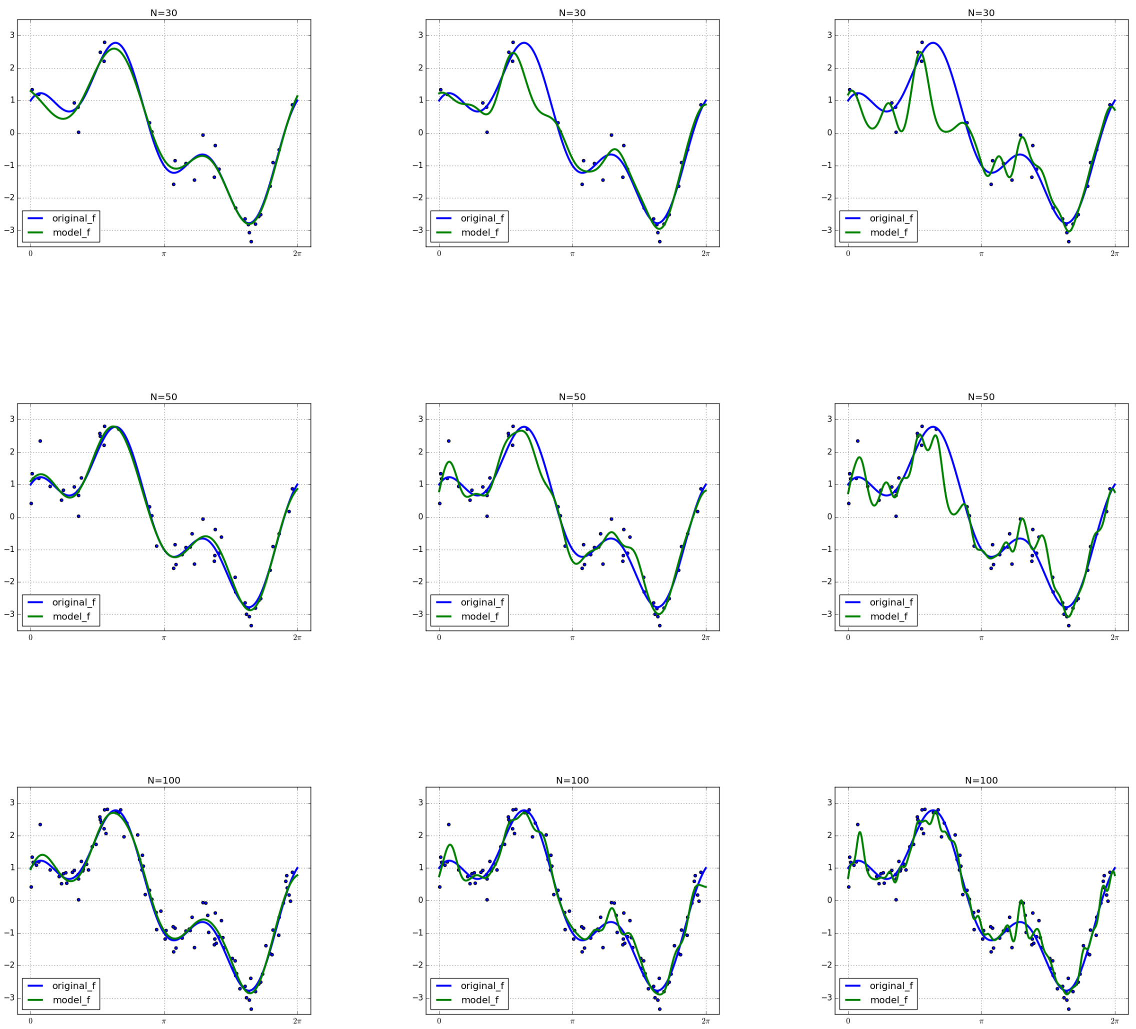
$$k(x_n, x_m) = exp\{-\frac{\theta}{2}(x_n - x_m)^2\}$$

を用いて, $\theta = 0.05 \times N, 0.3 \times N, 1 \times N$ としてデータ数 N$= 10, 30, 50, 100$ にそれぞれ試してみた









θ \ N	10	30	50	100
$0.05 \times N$	1.212	0.332	0.297	0.282
$0.3 \times N$	4.432	0.665	0.405	0.332
$1 \times N$	0.869	1.069	0.684	0.389

表 1: N と E,θ の関係

データ数によって最適なパラメータも変化すると考えられる．パラメータの最適化が重要になる．

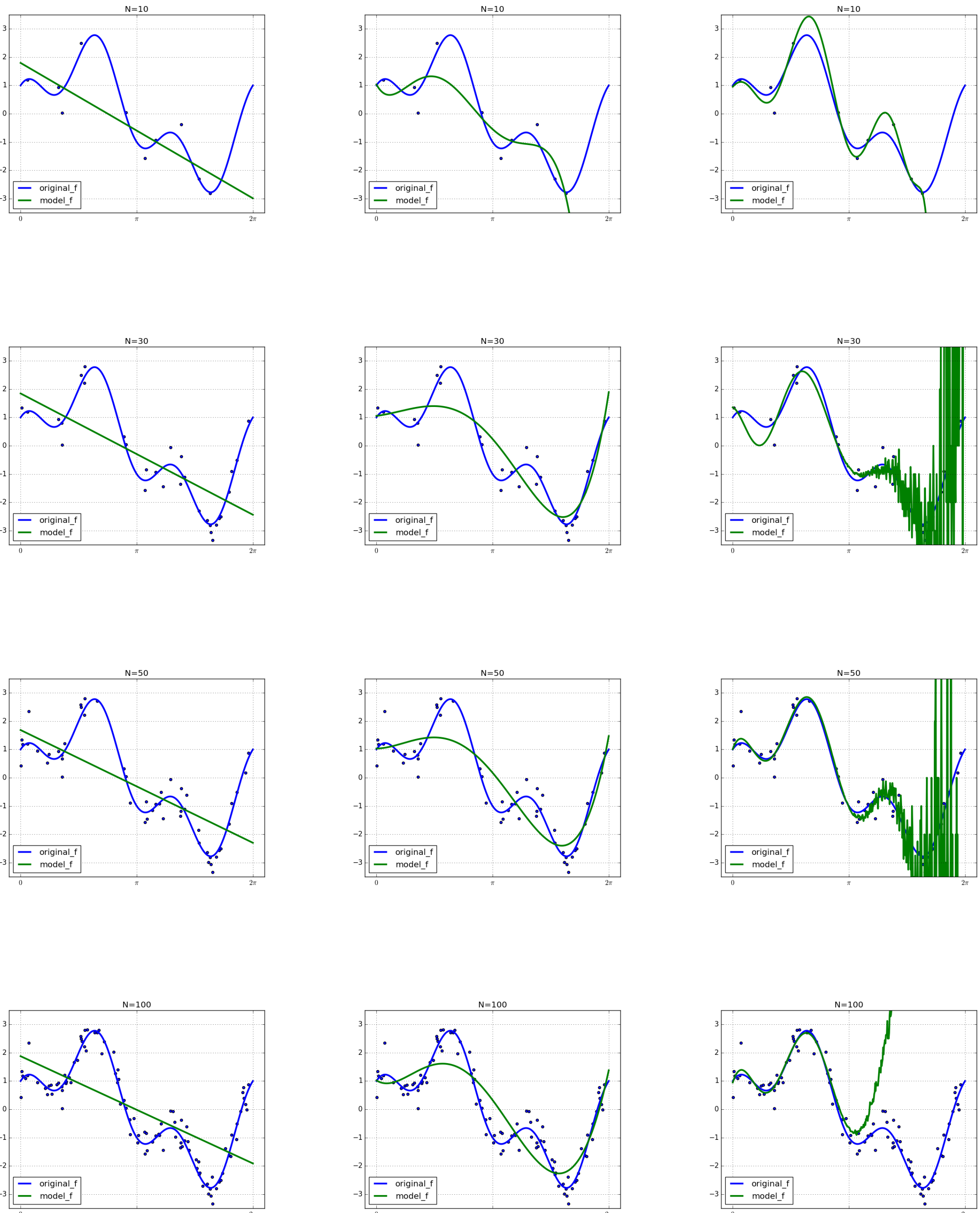
また，予測分布のパラメータは ($N = 100, \theta = 0.05$ で)$x = \pi$ に対し，平均 $= -0.925$(正解 $t = -1$), 分散 $= 11.126$(正解$\beta = 1/0.3^2$) となり両方とも正解に近くなっている．これは回帰が成功したものではおおよそ同じ結果となった．

4.2 多項式カーネル

ここでは，$\beta = 1/0.3^2$ としカーネル関数に

$$k(x_n, x_m) = (x_n x_m + 1)^{\theta}$$

を用いて, $\theta = 1, 5, 10$ としてデータ数 N$= 10, 30, 50, 100$ にそれぞれ試してみた



θ \ N	10	30	50	100
1	1.510	1.350	1.345	1.257
5	5.689	0.821	0.803	0.736
10	126.486	3.899	7.965	31.603

表 2: N と E,θ の関係

絶対値の大きな区間と小さな区間とで回帰の成功度合いが異なる．絶対値の大きな区間では回帰が難しくなってしまう．
また, 分散が正解と大きく異なる (分散が非常に大きい場合) 回帰がうまくいっていない傾向がみられる．

5 超パラメータの学習

$p(\mathbf{t}|\theta)$ は

$$p(\mathbf{t}|\theta) = N(\mathbf{t}|\mathbf{0}, C_N) \quad (6.61)'$$

となるため, この対数を θ について最大化する．

$$\ln p(\mathbf{y}|\theta) = -\frac{1}{2}|C_N| - \mathbf{t}^T C_N^{-1}\mathbf{t} - \frac{N}{2}\ln(2\pi) \quad (6.69)$$

また, θ はカーネル関数のパラメータなので, C_N に暗に含まれており, 上の対数の θ_i による微分は

$$\frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta_i} = -\frac{1}{2}Tr\left(C_N^{-1}\frac{\partial C_N}{\partial \theta_i}\right) + \frac{1}{2}\mathbf{t}^T C_N^{-1}\frac{\partial C_N}{\partial \theta_i}\mathbf{t} \quad (6.70)$$

となるが, $\ln p(\mathbf{t}|\theta)$ は一般に非凸関数なので複数の極大点を持つ．ただし完全にベイズ的に扱うには θ の事前分布との積を最大化する必要があり, これには近似を用いる．

6 関連度自動決定

入力が多次元であるとき, どの変数が学習にどの程度影響を与えるかを知ることができる．カーネル関数に

$$k(\mathbf{x}, \mathbf{z}) = \theta_0 exp - \frac{1}{2}\sum_i \eta_i (x_i - z_i)^2 \quad (6.71)$$

を用いて, 最尤推定からパラメータ η をデータに適応させる．このパラメータ η が, 学習への影響度合いを示す．

7 まとめ

ガウスカーネルは結構うまくいったが, 多項式カーネルはやはりうまくいかない．また, カーネル関数のパラメータを最適化する必要も大きいが, 結構式がややこしくなって難しそう．