

## 3.5 エビデンス近似

平成 28 年 9 月 11 日

### 概 要

PRML の「3.5 エビデンス近似」についての実装と考察

### 目 次

<b>1</b>	<b>問題設定</b>	<b>2</b>
<b>2</b>	<b>アルゴリズム</b>	<b>2</b>
2.1	$\alpha$ の最適化 . . . . .	3
2.2	$\beta$ の最適化 . . . . .	3
<b>3</b>	<b>コード</b>	<b>4</b>
<b>4</b>	<b>結果</b>	<b>5</b>
4.1	多項式基底 . . . . .	5
4.2	ガウス基底 . . . . .	6
4.3	シグモイド基底 . . . . .	7
4.4	tanh 基底 . . . . .	8
<b>5</b>	<b>まとめ</b>	<b>10</b>
<b>6</b>	<b><math>N \gg M</math> のとき</b>	<b>10</b>
6.1	アルゴリズム . . . . .	10
6.2	コード . . . . .	10
6.3	結果 . . . . .	10
6.3.1	多項式基底 . . . . .	10
6.3.2	ガウス基底 . . . . .	11
6.3.3	シグモイド基底 . . . . .	13
6.3.4	tanh 基底 . . . . .	14
6.4	まとめ . . . . .	15
<b>7</b>	<b>追記</b>	<b>15</b>

## 1 問題設定

線形基底関数モデルを完全にベイズ的に扱うために、超パラメータ  $\alpha, \beta$  に対しても事前分布を導入し、通常のパラメータ  $\mathbf{w}$  だけでなく、超パラメータに関しても周辺歌詞予測を行う。パラメータ  $\mathbf{W}$  だけに関して積分して得られた周辺尤度関数を最大にするように超パラメータの値を決めるという2段階の近似法がエビデンス近似である。

## 2 アルゴリズム

これは、

$$p(t|\mathbf{t}) = \int \int \int p(t|\mathbf{w}, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta) p(\alpha, \beta|\mathbf{t}) d\mathbf{w} d\alpha d\beta \quad (3.74)$$

となるものを、

$$p(t|\mathbf{t}) = \int p(t|\mathbf{w}, \hat{\beta}) p(\mathbf{w}|\mathbf{t}, \hat{\alpha}, \hat{\beta}) d\mathbf{w} \quad (3.75)$$

と近似するのである。ここで、 $\hat{\alpha}, \hat{\beta}$  は事後分布  $p(\alpha, \beta|\mathbf{t})$  を最大化するものとして与えられる。ここで、 $\alpha, \beta$  の事後分布  $p(\alpha, \beta|\mathbf{t})$  は、ベイズの定理より、

$$p(\alpha, \beta|\mathbf{t}) \propto p(\mathbf{t}|\alpha, \beta) p(\alpha, \beta) \quad (3.76)$$

となる。エビデンス近似の枠組みでは、事前分布  $p(\alpha, \beta)$  は平坦であると仮定し、 $p(\mathbf{t}|\alpha, \beta)$  の最大化について考える。

$$p(\mathbf{t}|\alpha, \beta) = \int p(\mathbf{t}|\mathbf{w}, \beta) p(\mathbf{w}|\alpha) d\mathbf{w} \quad (3.77)$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \left(\frac{\alpha}{2\pi}\right)^{M/2} \int \exp\{-E(\mathbf{w})\} d\mathbf{w} \quad (3.78)$$

ここで、

$$\begin{aligned} E(\mathbf{w}) &= \beta E_D(\mathbf{w}) + \alpha E_W(\mathbf{w}) \\ &= \frac{\beta}{2} \|\mathbf{t} - \Phi \mathbf{w}\|^2 + \frac{\alpha}{2} \mathbf{w}^T \mathbf{w} \end{aligned} \quad (3.79)$$

また、 $\mathbf{w}$  に関して平方完成することで、

$$E(\mathbf{w}) = E(\mathbf{m}_N) + \frac{1}{2} (\mathbf{w} - \mathbf{m}_N)^T A (\mathbf{w} - \mathbf{m}_N) \quad (3.80)$$

ただし、

$$A = \alpha I + \beta \Phi^T \Phi \quad (3.81) \quad \mathbf{m}_N = \beta A^{-1} \Phi^T \mathbf{t} \quad (3.84)$$

と定義した。

周辺尤度について計算する、

$$\begin{aligned} \int \exp\{-E(\mathbf{w})\} d\mathbf{w} &= \exp\{-E(\mathbf{m}_N)\} \int \exp\left\{-\frac{1}{2} (\mathbf{w} - \mathbf{m}_N)^T A (\mathbf{w} - \mathbf{m}_N)\right\} d\mathbf{w} \\ &= \exp\{-E(\mathbf{m}_N)\} (2\pi)^{M/2} |A|^{-1/2} \end{aligned} \quad (3.85)$$

そして、(3.78) の対数つまり、対数尤度は、

$$\ln p(\mathbf{t}|\alpha, \beta) = \frac{M}{2} \ln \alpha + \frac{N}{2} \ln \beta - E(\mathbf{m}_N) - \frac{1}{2} \ln |A| - \frac{N}{2} \ln(2\pi) \quad (3.86)$$

## 2.1 $\alpha$ の最適化

まず, 固有方程式,

$$(\beta \Phi^T \Phi) \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (3.87)$$

が成り立つとすると,  $A$  は固有値  $\alpha + \lambda_i$  を持つ. これより, (3.86) の  $\alpha$  による微分について,

$$\frac{d}{d\alpha} \ln |A| = \frac{d}{d\alpha} \ln \prod_i (\lambda_i + \alpha) = \frac{d}{d\alpha} \sum_i \ln(\lambda_i + \alpha) = \sum_i \frac{1}{\lambda_i + \alpha} \quad (3.88)$$

となるので,

$$0 = \frac{M}{2\alpha} - \frac{1}{2} \mathbf{m}_N^T \mathbf{m}_N - \frac{1}{2} \sum_i \frac{1}{\lambda_i + \alpha} \quad (3.89)$$

これより,

$$\alpha \mathbf{m}_N^T \mathbf{m}_N = M - \alpha \sum_i \frac{1}{\lambda_i + \alpha} = \gamma \quad (3.90)$$

第2項を整理すると,

$$\gamma = \sum_i \frac{1}{\lambda_i + \alpha} \quad (3.91)$$

(3.90) より,

$$\alpha = \frac{\gamma}{\mathbf{m}_N^T \mathbf{m}_N} \quad (3.92)$$

となる.

エビデンス近似

1.  $\Phi^T \Phi$  の固有値を計算する.
2. 1で求めた固有値に  $\beta$  を掛けたものが  $\lambda$  で, これと現在の  $\alpha$  から (3.91) を用いて  $\gamma$  を計算する.
3. また, 現在の  $\alpha$  から  $\mathbf{m}_N$  を計算する.
4. (3.92) から  $\alpha$  を計算し, 更新する. 更新前後の  $\alpha$  の差が十分小さくなるまで2に戻り繰り返す.

## 2.2 $\beta$ の最適化

同様に, (3.86) を  $\beta$  で微分すると

$$\frac{d}{d\beta} \ln |A| = \frac{d}{d\beta} \sum_i \ln(\lambda_i + \alpha) = \frac{1}{\beta} \sum_i \frac{\lambda_i}{\lambda_i + \alpha} = \frac{\gamma}{\beta} \quad (3.93)$$

となるため,

$$0 = \frac{N}{2\beta} - \frac{1}{2} \sum_{n=1}^N \{t_n - \mathbf{m}_N \phi(x_n)\}^2 - \frac{\gamma}{2\beta} \quad (3.94)$$

よって,

$$\frac{1}{\beta} = \frac{1}{N - \gamma} \sum_{n=1}^N \{t_n - \mathbf{m}_N \phi(x_n)\}^2 \quad (3.95)$$

となる.

#### エビデンス近似

1.  $\Phi^T \Phi$  の固有値を計算する.
2. 1 で求めた固有値に現在の  $\beta$  を掛けたものが  $\lambda$  で, これと  $\alpha$  から (3.91) を用いて  $\gamma$  を計算する.
3. また,  $\alpha$  から  $\mathbf{m}_N$  を計算する.
4. (3.95) から  $\beta$  を計算し, 更新する. 更新前後の  $\beta$  の差が十分小さくなるまで 2 に戻り繰り返す.

### 3 コード

```
for N in [20,100,500]:
    x=data[:N,0]
    t=data[:N,1]
    for M in [4,10,20]:
        """ W の最適化 """
        A=np.zeros((M,M))
        W=np.zeros(M)
        P=np.zeros((N,M))
        I=np.identity(M)

        mu=[(m+0.5)*(2*pi/(M+1)) for m in range(M)]
        s=(2*pi)**2/12

        for n in range(N):
            for m in range(M):
                P[n,m]=tanh_basis(x[n],m,mu[m],s)

        alpha,beta=0,1.0/(0.3)**2
        delta1,delta2=1,1

        r=eigvals(np.dot(P.T,P))

        while delta1>10**(-10) or delta2>10**(-10):
            mae_alpha,mae_beta=alpha,beta
            S_N=inv(mae_alpha*I+mae_beta*np.dot(P.T,P))
            m_N=mae_beta*np.dot(S_N,np.dot(P.T,t))
            new_r=r*mae_beta
            lam=0
            for m in range(M):
                lam+=new_r[m]/(mae_alpha+new_r[m])
            alpha=lam/np.dot(m_N,m_N)
            waru=0
            for n in range(N):
                waru+=(t[n]-np.dot(m_N.T,P[n,:]))**2
            beta=(N-lam)/waru
            delta1,delta2=alpha-mae_alpha,beta-mae_beta

        #求まったパラメータからモデル関数を作り
        def model_f(x):
            sum=0
            for m in range(M):
```

```

        sum+=m_N[m]*tanh_basis(x,m,mu[m],s)
    return sum

"""表示"""
#誤差関数を定義し出力
def Error():
    sum=0
    for n in range(50):
        sum+=(model_f(new_x[n])-new_t[n])**2
    return np.sqrt(sum/50)

print(N,"\t",M,"\t%.2f\t%3e\t%3e" % (Error(),alpha,beta))

```

## 4 結果

### 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。

M \ N	20	100	500
4	0.99	0.79	0.77
10	1.08	0.43	0.38
20	nan	6.77	599

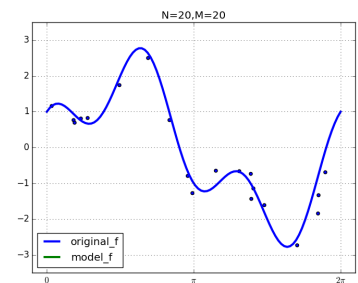
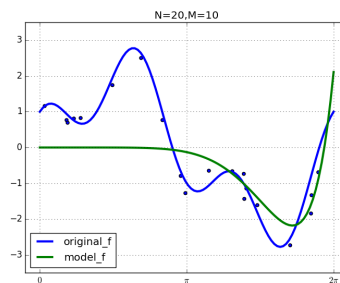
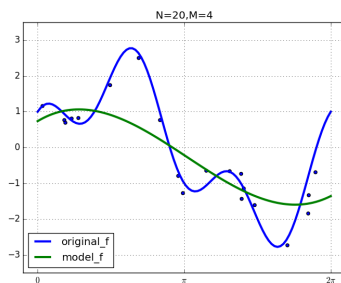
表 1:  $E_{RMS}$  の N,M との関係 (多項式基底)

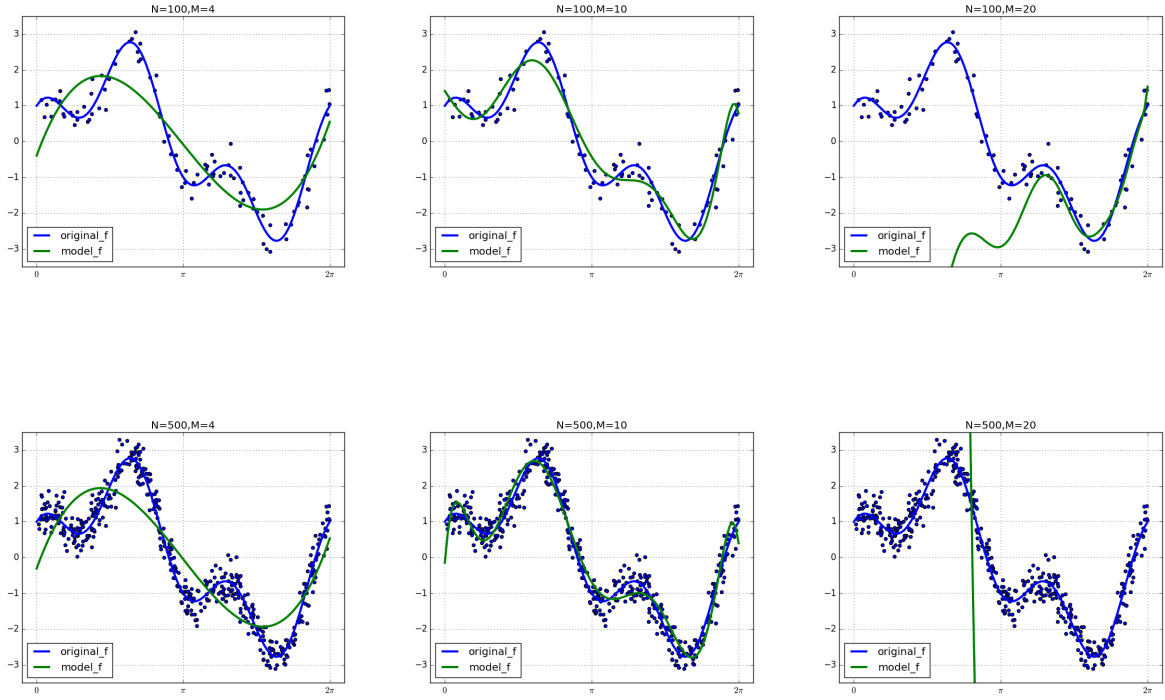
M \ N	20	100	500
4	2.2e+00	2.5e-01	2.5e-01
10	2.8e+09	4.0e-01	1.1e-03
20	nan	3.5e-02	9.8e-07

表 2:  $\alpha$

M \ N	20	100	500
4	1.8e+00	1.8e+00	1.5e+00
10	1.1e+00	4.3e+00	6.6e+00
20	nan	1.2e-02	2.3e-06

表 3:  $\beta$





## 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  $\mu$  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.

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

M \ N	N		
	20	100	500
4	1.08	1.01	0.98
10	1.30	0.61	0.90
20	0.94	1.27	3.75

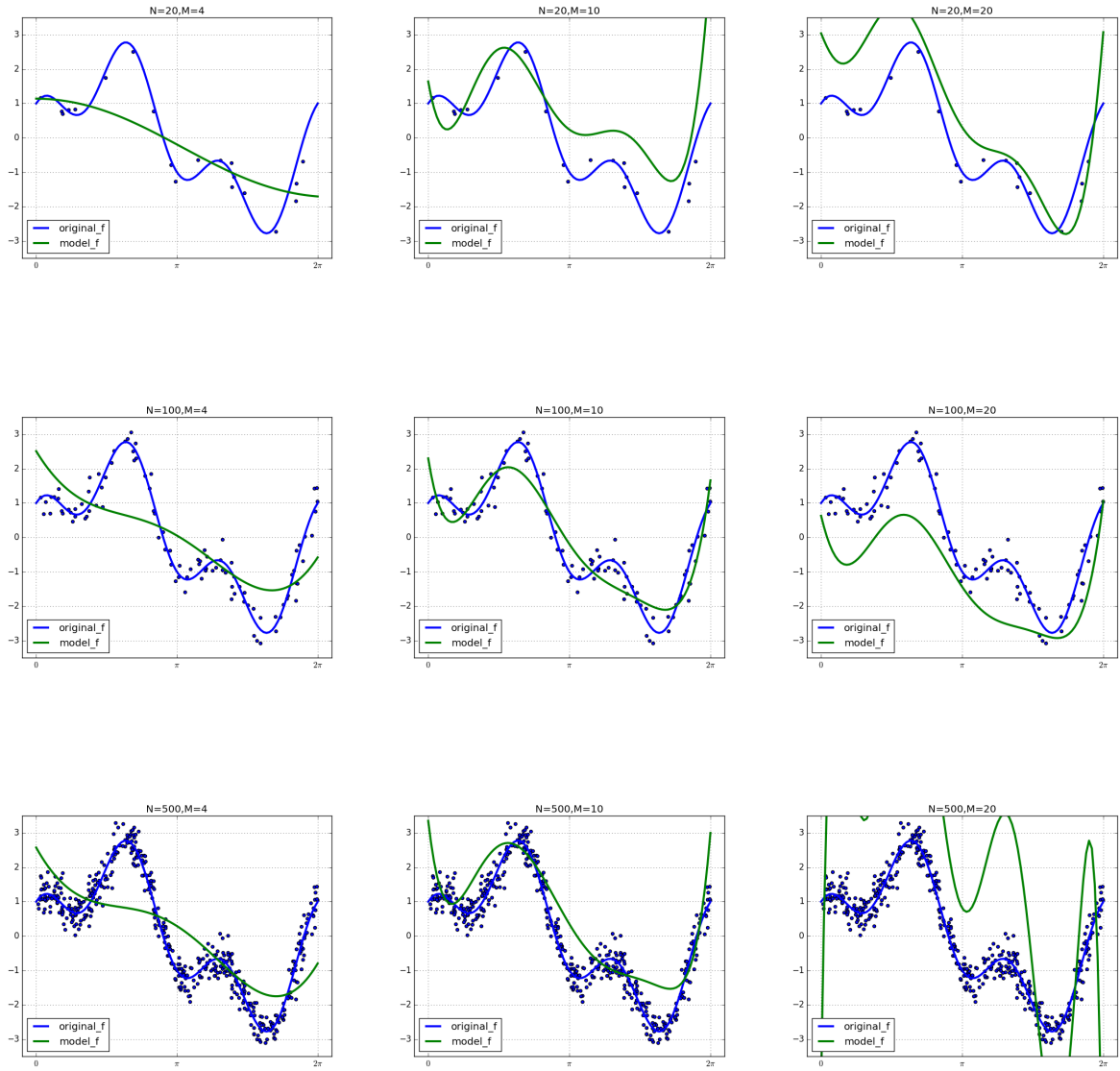
表 4:  $E_{RMS}$  の  $N, M$  との関係 (多項式基底)

M \ N	N		
	20	100	500
4	1.2e-01	4.1e-05	3.0e-05
10	1.3e-14	4.6e-13	1.8e-14
20	1.9e-14	4.6e-14	6.8e-17

表 5:  $\alpha$

M \ N	N		
	20	100	500
4	1.6e+00	9.9e-01	9.2e-01
10	4.8e-01	2.8e+00	1.5e+00
20	0.0e+00	4.6e-01	6.2e-02

表 6:  $\beta$



### 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	N		
	20	100	500
4	1.15	0.79	0.77
10	2.55	0.77	0.61
20	1.34	0.88	0.66

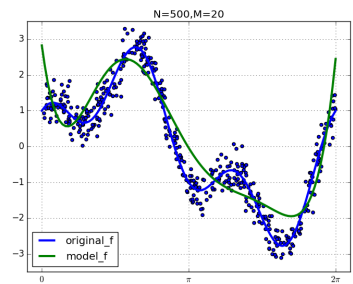
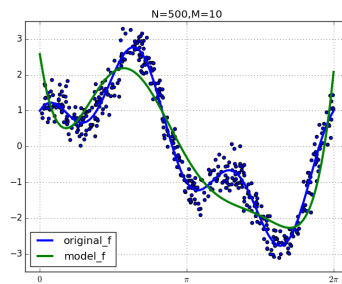
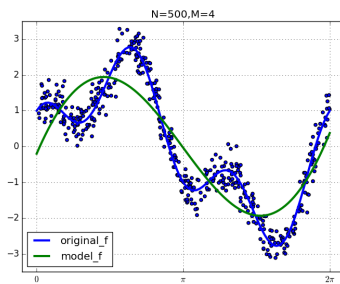
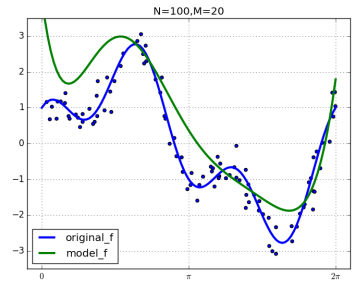
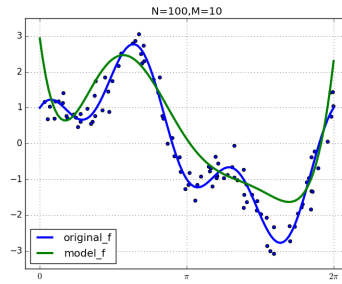
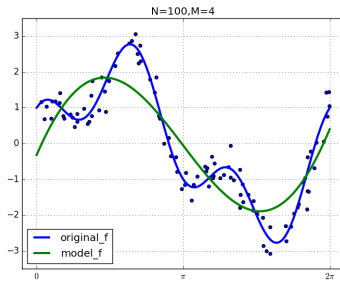
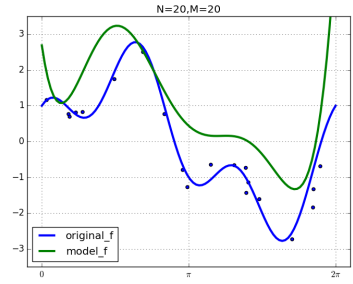
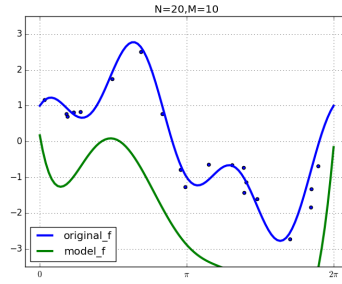
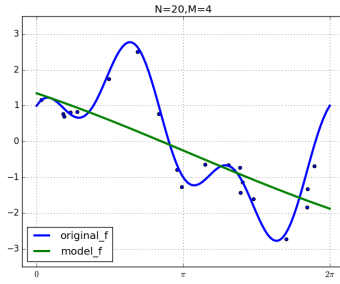
表 7:  $E_{RMS}$  の N,M との関係 (多項式基底)

M \ N	20	100	500
4	6.5e-02	2.2e-07	2.1e-07
10	3.5e-16	8.5e-15	1.2e-14
20	4.2e-14	6.0e-14	2.7e-13

表 8:  $\alpha$

M \ N	20	100	500
4	1.5e+00	1.8e+00	1.6e+00
10	8.5e-02	2.0e+00	2.6e+00
20	0.0e+00	9.9e-01	2.5e+00

表 9:  $\beta$



#### 4.4 tanh 基底

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



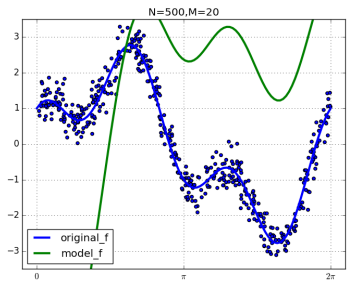
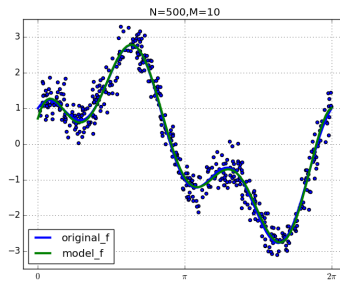
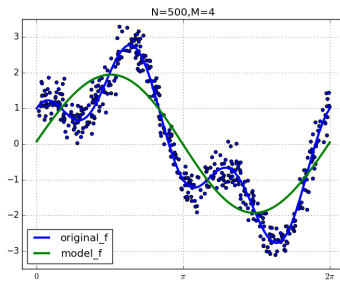
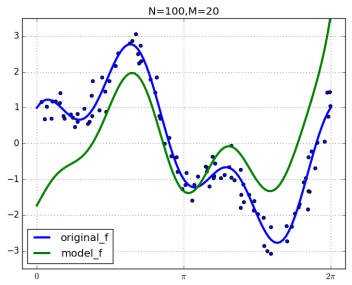
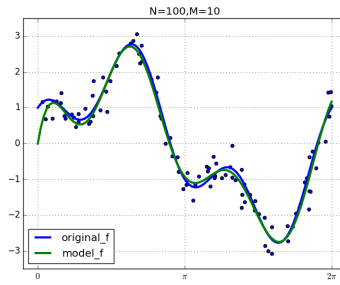
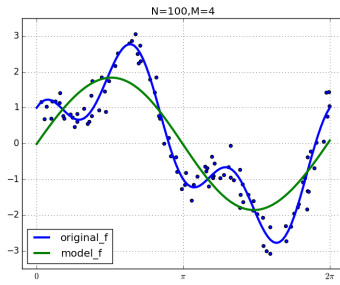
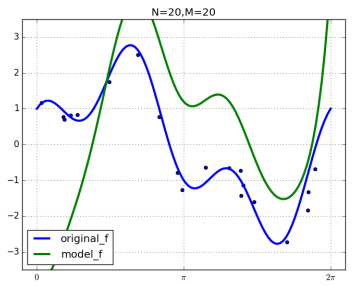
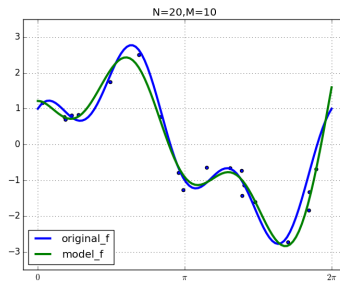
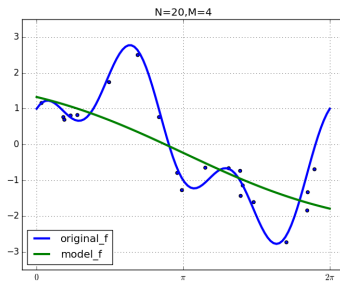
M \ N	N		
	20	100	500
4	1.12	0.79	0.78
10	0.39	0.33	0.28
20	2.19	1.26	4.80

表 10:  $E_{RMS}$  の N,M との関係 (多項式基底)

M \ N	N			M \ N	N		
	20	100	500		20	100	500
4	1.2e+00	3.2e-04	3.1e-04	4	1.6e+00	1.8e+00	1.6e+00
10	3.9e-11	1.9e-12	1.2e-11	10	1.2e+01	8.3e+00	9.6e+00
20	8.0e-14	2.3e-13	1.1e-12	20	0.0e+00	4.7e-01	3.8e-02

表 11:  $\alpha$

表 12:  $\beta$



## 5 まとめ

$N = 500, M = 20$  でおかしい挙動を起こすものが多い. また, 多項式基底ではあまりうまくいかない.

結局  $M = 10$  程度のときが一番学習精度が良かった気がする. モデルサイズは, 十分かつ適当であるべき.

## 6 $N \gg M$ のとき

### 6.1 アルゴリズム

全てのパラメータは well-defined となり,  $\gamma \rightarrow M$  と扱う. この結果,

$$\alpha = \frac{M}{2E_W(\mathbf{m}_N)} \quad (3.98) \quad \beta = \frac{N}{2E_D(\mathbf{m}_N)} \quad (3.99)$$

となる. これを, 再定義式に用いる.

### 6.2 コード

アルゴリズムの本体 (test1.py)

```
alpha,beta=0,1.0/(0.3)**2
delta1,delta2=1,1

while delta1>10**(-10) or delta2>10**(-4):
    mae_alpha,mae_beta=alpha,beta
    S_N=inv(mae_alpha*I+mae_beta*np.dot(P.T,P))
    m_N=mae_beta*np.dot(S_N,np.dot(P.T,t))
    alpha=M/np.dot(m_N,m_N)
    waru=0
    for n in range(N):
        waru+=(t[n]-np.dot(m_N,P[n,:]))**2
    beta=N/waru
    delta1,delta2=alpha-mae_alpha,beta-mae_beta
```

### 6.3 結果

#### 6.3.1 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.

$M \backslash N$	20	100	500
4	1.01	0.79	0.77
10	nan	0.43	0.38
20	58.4	36.2	11.4

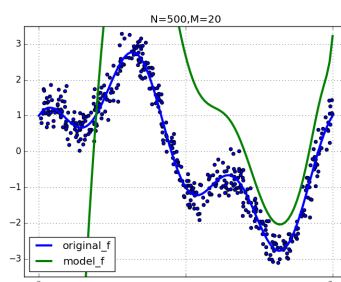
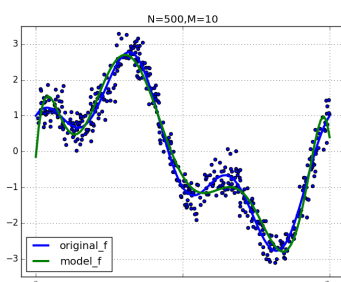
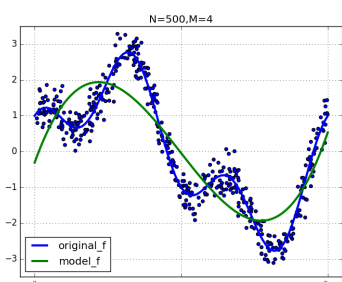
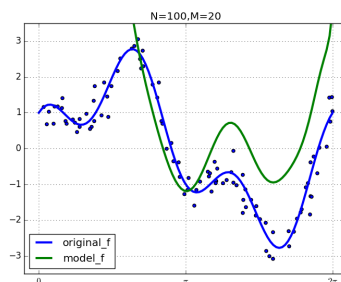
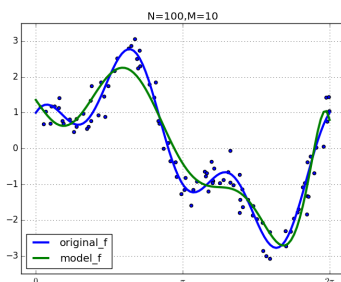
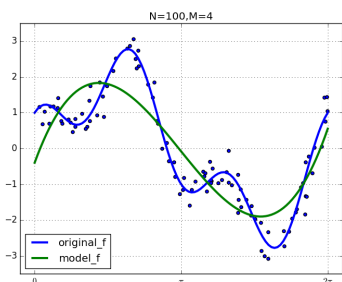
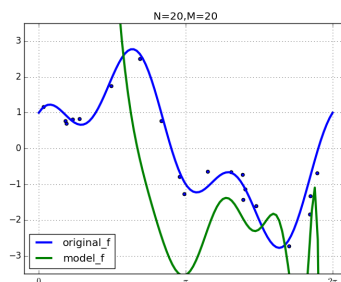
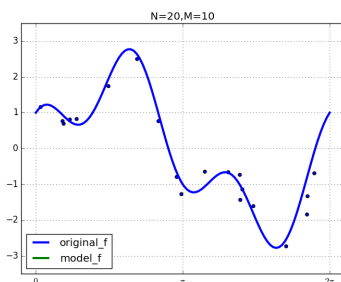
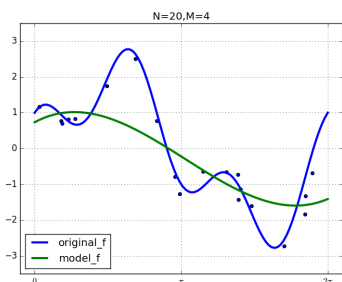
表 13:  $E_{RMS}$  の  $N, M$  との関係 (多項式基底)

M \ N	20	100	500
4	3.2e+00	2.5e-01	2.5e-01
10	nan	5.5e-01	1.1e-03
20	5.4e-05	1.9e-04	1.6e-03

表 14:  $\alpha$

M \ N	20	100	500
4	2.1e+00	1.9e+00	1.5e+00
10	nan	4.6e+00	6.8e+00
20	3.0e-04	8.3e-04	8.0e-03

表 15:  $\beta$



### 6.3.2 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  $\mu$  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.

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

M \ N	N		
	20	100	500
4	1.09	1.11	0.98
10	1.30	0.61	0.90
20	1.27	1.13	3.75

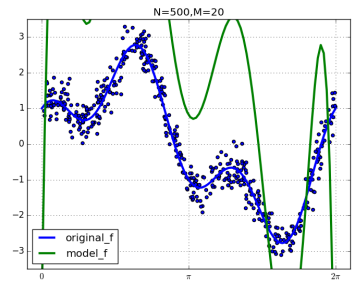
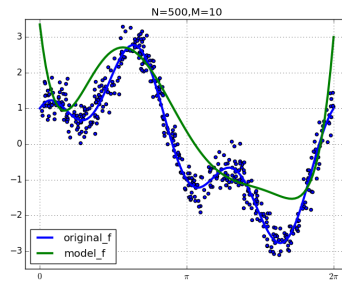
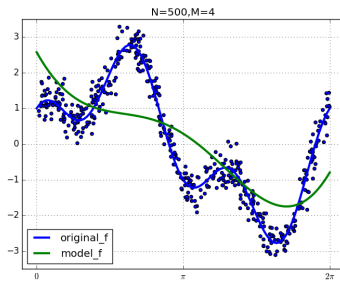
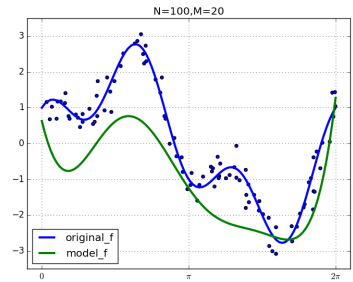
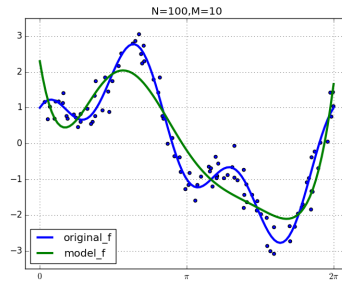
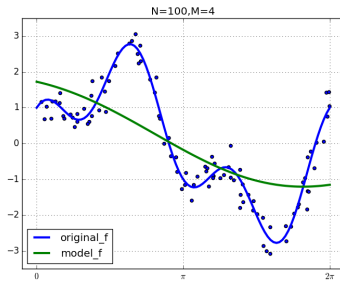
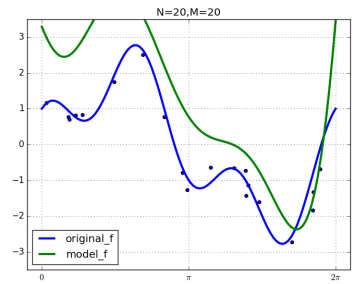
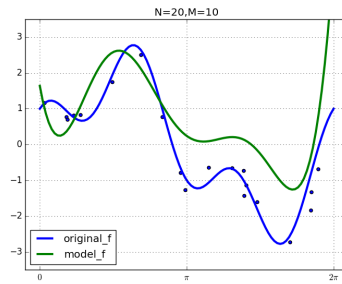
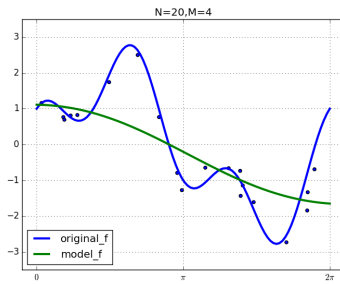
表 16:  $E_{RMS}$  の  $N, M$  との関係 (多項式基底)

M \ N	N		
	20	100	500
4	2.0e-01	1.7e-01	3.1e-05
10	1.3e-14	4.6e-13	1.8e-14
20	1.9e-14	4.6e-14	6.8e-17

表 17:  $\alpha$

M \ N	N		
	20	100	500
4	1.8e+00	9.1e-01	9.3e-01
10	9.7e-01	3.1e+00	1.5e+00
20	5.2e-01	7.0e-01	6.5e-02

表 18:  $\beta$



### 6.3.3 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	1.17	0.79	0.77
10	2.55	0.77	0.61
20	1.34	0.80	0.66

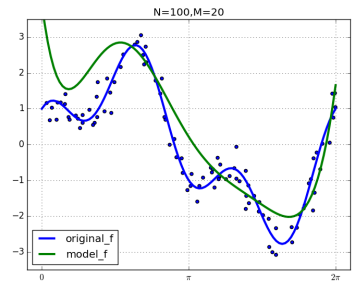
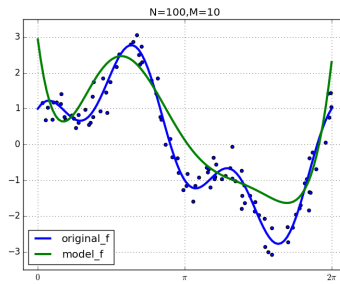
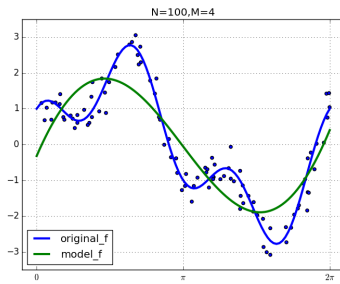
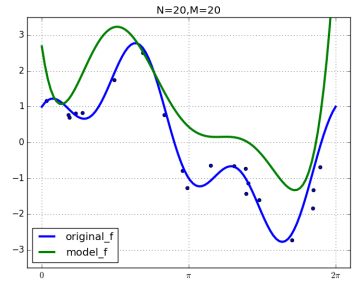
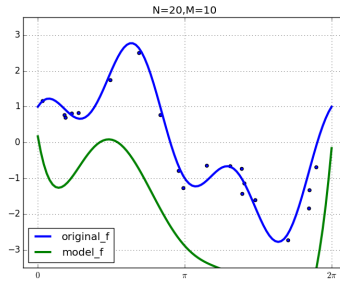
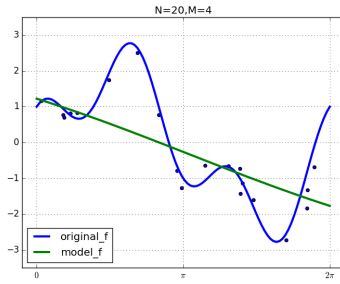
表 19:  $E_{RMS}$  の N,M との関係 (多項式基底)

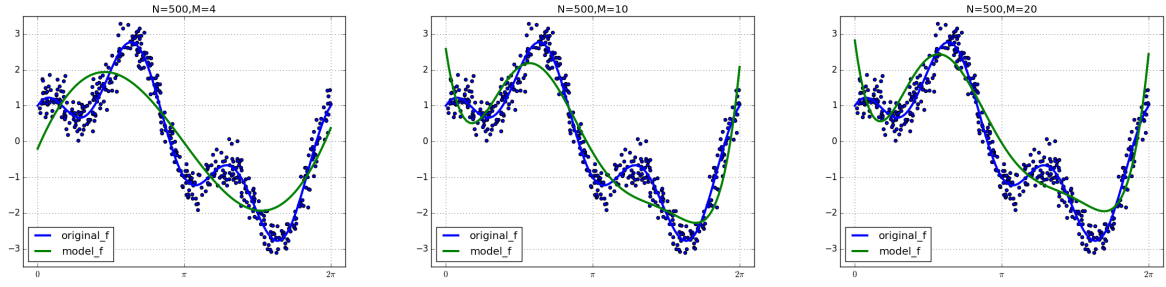
M \ N	20	100	500
4	1.6e-01	2.2e-07	2.1e-07
10	3.5e-16	8.5e-15	1.2e-14
20	4.2e-14	6.0e-14	2.7e-13

表 20:  $\alpha$

M \ N	20	100	500
4	1.6e+00	1.9e+00	1.6e+00
10	1.7e-01	2.2e+00	2.6e+00
20	8.2e-01	1.5e+00	2.6e+00

表 21:  $\beta$





### 6.3.4 tanh 基底

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

M \ N	N		
	20	100	500
4	1.14	0.79	0.78
10	0.35	0.33	0.28
20	2.19	1.32	4.80

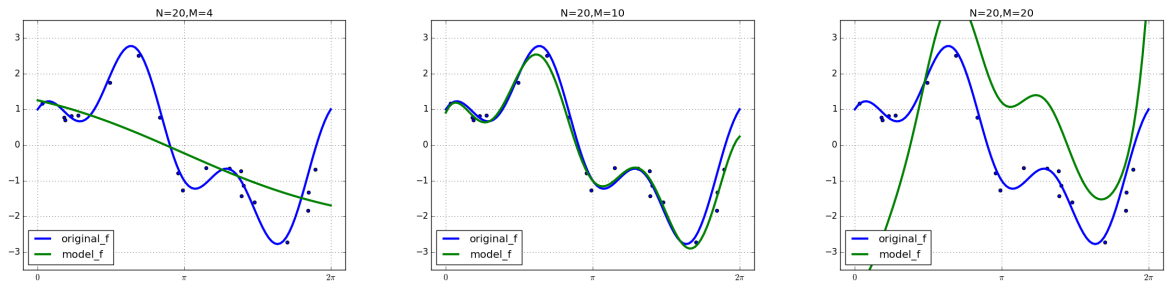
表 22:  $E_{RMS}$  の N,M との関係 (多項式基底)

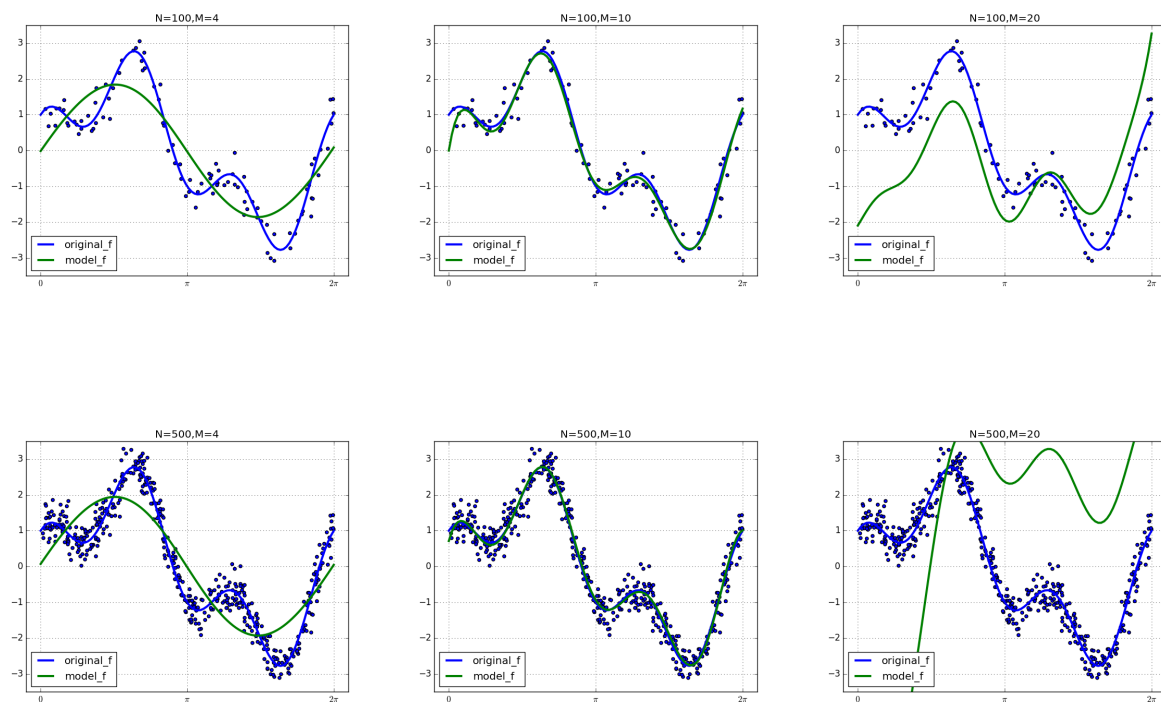
M \ N	N		
	20	100	500
4	2.7e+00	3.3e-04	3.1e-04
10	1.5e-11	1.9e-12	1.2e-11
20	8.0e-14	2.3e-13	1.1e-12

表 23:  $\alpha$

M \ N	N		
	20	100	500
4	1.7e+00	1.9e+00	1.6e+00
10	2.5e+01	9.2e+00	9.7e+00
20	1.9e-01	5.0e-01	4.0e-02

表 24:  $\beta$





## 6.4 まとめ

あらい近似に見えたが, そうでもなかった. あまり, 完全なエビデンス近似との差はなかった.

## 7 追記

2 アルゴリズムでは, 事前分布  $p(\alpha, \beta)$  は平坦であると仮定しているが, ここで共役事前分布としてガンマ分布を用いて解析することも可能で, 事後分布は, スチューデントの  $t$  分布になる. しかし, 近似を用いて解析するしかなくなり, あまり近似精度もよくない.