

カーネル回帰関数の確率変数化による ノイズを含む入力に対する回帰手法

Regression for Noisy Input using a Kernel Regression Function for Noiseless Input

花房 諒 岡留 剛
Ryo Hanafusa Takeshi Okadome

関西学院大学大学院理工学研究科
Graduate School of Science and Technology, Kwansei Gakuin University

The regression method proposed here determines the regression function for noisy inputs that minimizes the mean squared error of the prediction with the error in the objective variable. The method begins with the kernel regression function $\hat{t} = y(\mathbf{x})$ built from the training dataset that consists of value pairs of a noiseless explanatory variable and an objective variable. Given a noisy input $\mathbf{u} = \mathbf{x} + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \sim p_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{\kappa})$ is noise, the method finds the posterior of the noiseless input $p(\mathbf{x}|\mathbf{u})$ by the Bayes' theorem. Then, the method obtains the predictive distribution $p(\hat{t}|\mathbf{u})$ using the kernel regression function $\hat{t} = y(\mathbf{x})$ with the posterior $p(\mathbf{x}|\mathbf{u})$. The regression value for the input $\mathbf{u}$ is the expectation of the predictive distribution. The method suppresses the overfitting of the regression function for noisy inputs and the RMSEs of the predictions are lower compared to those of an exiting method.

1. はじめに

本研究では、ノイズを含む入力に対する回帰手法を提案する。その手法は、カーネル密度推定を利用したノイズなし入力に対する回帰関数である Nadaraya-Watson カーネル回帰関数を基礎とする。その上で、目的変数の誤差を考慮し、目的変数とノイズを含む入力に対する回帰との期待二乗損失を最小にする回帰関数を与える。

本研究で扱う問題は、説明変数にノイズがともなう回帰問題として、以下の2つを前提とする。(a) 訓練時には、ノイズを含む説明変数値とそれに対応する目的変数値は得にくく、ノイズを含まない説明変数値とそれに対応する目的変数値の組が与えられる。(b) 実際に回帰関数を適用する場合には、新たな入力にノイズが含まれる。例えば、実験室で得られるようなデータから構築される回帰関数に対し、実験室の外で得られる新たな入力にはノイズが含まれる、という状況である。また、前提 a により、ノイズなし入力に対する回帰関数は構築できる。

ノイズを仮定せずに学習した回帰関数にノイズが含まれた入力を与えた場合、それにより得られる目的変数の推定値は誤差が大きい。そのため、現実では通常、新たなノイズあり入力からノイズを取り除いて回帰関数への入力とする。しかし、その入力からノイズを取り除く時間がない状況や、ノイズを取り除く方法がない場合がある。医療分野でのその具体例として、健康診断と緊急検査が挙げられる。健康診断では、食事や薬などの生理的変動因子の影響がノイズとなるため、これを取り除いてから患者の検査を行なう。すなわち、健康診断で得られる検査値はノイズはなく、その値からさまざまな病気に対する危険度が判断される。一方、緊急検査ではノイズを取り除かず急患に対して検査を行なう。その患者は、検査前に食事や薬の服用をしている可能性があり、得られる検査値にはノイズが含まれている可能性が高い [濱崎 08][北村 14]。そのような場合、健康診断時の検査値とその検査値に対する病気の危険度を利用し、緊急検査時の危険度を推定することができる。一般に、病気の危険度を判断するための基準値には個人差があることが知られている [Kamatani 10][下方 13]。個人差を考慮した場合、

病気の危険度は医師により決定されるため、医師の違いによって誤差が発生する。本研究では、このような状況での回帰問題を扱う。すなわち、訓練時の説明変数値にはノイズが含まれないが、テスト時には入力にノイズが含まれ、さらに、訓練時の目的変数値には誤差が含まれる。

ノイズを含む入力に対する回帰として Webb の手法がある [Webb 94]。その手法は、目的変数とノイズあり入力に対する回帰との期待二乗損失を最小にする回帰関数を構築する。しかし、上記の例のような場合には Webb の手法を適用することは適切でない。Webb の手法では、訓練データ集合の目的変数値に誤差が含まれることを考えていない。そのため、新たな入力に対する回帰が目的変数値の誤差に追従してしまい、回帰の精度が落ちる可能性がある。また、入力ノイズが従う真の分布が分かっているか、あるいは、その分布を適当に仮定しなければならない。現実では、真の入力ノイズ分布が分かっているような状況は考えにくく、また、適当に仮定した入力ノイズ分布を用いた回帰の結果は信用性が低い。そのため、目的変数の誤差を考慮し、目的変数とノイズあり入力に対する回帰との期待二乗損失を最小にする回帰手法を提案する。これにより、Webb の手法で求まる回帰関数で起こる過学習が、本提案手法による回帰関数では起こらない。

本提案手法は、以下の3つの手順からなる。(1) ノイズを含まない説明変数値と目的変数値の組からなる訓練データ集合から構築される、ノイズなし入力に対する回帰関数 $\hat{t} = y(\mathbf{x})$ を出発点とする。(2) ノイズ $\boldsymbol{\eta}$ を含む入力 $\mathbf{u} = \mathbf{x} + \boldsymbol{\eta}$ と入力ノイズ分布 $p_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{\kappa})$ が与えられたもとの、ベイズの定理によりノイズを除いた入力の事後分布 $p(\mathbf{x}|\mathbf{u})$ を求める。(3) その事後分布を用いて1の回帰関数を確率変数化することにより、ノイズを含む入力に対する回帰の予測分布 $p(\hat{t}|\mathbf{u})$ を求め、その分布の期待値 $E[\hat{t}|\mathbf{u}]$ を回帰の出力とする。

さらに、医療分野等への応用を目的に、時系列性をもつデータへの提案手法の適用も行なう。そのために、訓練データ集合の説明変数値とノイズを除いた入力を隠れマルコフモデル (以下、HMM) で表現する。また、入力ノイズ分布のパラメータの真の値は分からない状況を仮定し、パラメータ値の推定を行なう。パラメータ推定では、パラメータを確率変数とし事後確率最大化によって決定するアプローチをとる。

連絡先: 花房 諒, 関西学院大学大学院理工学研究科,
aoa93149@kwansei.ac.jp

2. 問題の定式化

ノイズあり入力を $\mathbf{u} = \mathbf{x} + \boldsymbol{\eta}$ とし、ノイズを除いた入力を $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\lambda})$ 、ノイズを $\boldsymbol{\eta} \sim p_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{\kappa})$ とする。変数 $\mathbf{u}, \mathbf{x}, \boldsymbol{\eta}$ は $\mathbb{R}^d$ に値をとる確率変数である。ただし、 $\boldsymbol{\lambda}, \boldsymbol{\kappa}$ は分布のパラメータであり、 $p_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{\kappa})$ は与えられるとし、 $p(\mathbf{x}|\boldsymbol{\lambda})$ は未知とする。ノイズを除いた入力 $\mathbf{x}$ とノイズ $\boldsymbol{\eta}$ は独立であると仮定する。それにより、入力ノイズ分布 $p_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{\kappa})$ は

$$p_{\boldsymbol{\eta}}(\mathbf{u} - \mathbf{x}|\boldsymbol{\kappa}) = p(\mathbf{u}|\mathbf{x}, \boldsymbol{\kappa}) \quad (1)$$

と表わすことができる。また、ノイズなし説明変数値 $\mathbf{x}_n$ と誤差を含む目的変数値 t_n の組からなる訓練データ集合を $D = \{(\mathbf{x}_n, t_n) | \mathbf{x}_n \in \mathbb{R}^d, t_n \in \mathbb{R}\}_{n=1, \dots, N}$ とする。訓練データ集合 D の説明変数値 $\mathbf{x}_n$, $n = 1, \dots, N$, とノイズを除いた入力 $\mathbf{x}$ は独立同分布に従うとする。

ノイズあり入力 $\mathbf{u}$ と入力ノイズの分布 $p_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{\kappa})$ 、訓練データ集合 D の3つが与えられたもとで、ノイズを除いた入力 $\mathbf{x}$ に対する目的変数 t を推定する。入力ノイズ分布のパラメータ $\boldsymbol{\kappa}$ は、真の値が与えられているときにはその値を用いるが、与えられないときは適当に値を仮定する。

3. 関連研究

ノイズあり入力に対する回帰として Webb の手法がある。この手法では、目的変数とノイズあり入力に対する回帰との期待二乗損失を最小にする回帰関数を構築する。その関数は

$$\hat{g}(\mathbf{u}) = \frac{\sum_{n=1}^N p_{\boldsymbol{\eta}}(\mathbf{u} - \mathbf{x}_n) t_n}{\sum_{m=1}^N p_{\boldsymbol{\eta}}(\mathbf{u} - \mathbf{x}_m)} \quad (2)$$

で与えられる [Webb 94][Bishop 95]。この手法では、目的変数の誤差は考慮されていない。式 (2) は、Nadaraya-Watson カーネル回帰関数の特殊な場合となっている*1。

4. 提案手法

4.1 ノイズを含む入力に対する回帰手法

本提案手法は、目的変数の誤差を考慮した上で、目的変数とノイズあり入力に対する回帰との期待二乗損失を最小にする回帰関数を与える。その基礎は、ノイズなし入力 $\mathbf{x}$ に対する回帰関数として用いる Nadaraya-Watson カーネル回帰関数である。そのカーネル回帰関数は、カーネル密度推定により得られる説明変数と目的変数との同時確率を利用する。

本手法の出発点として、ノイズなし入力 $\mathbf{x}$ に対する回帰関数である Nadaraya-Watson カーネル回帰関数

$$\hat{g}(\mathbf{x}) = \frac{\sum_{n=1}^N k(\mathbf{x}, \mathbf{x}_n) t_n}{\sum_{m=1}^N k(\mathbf{x}, \mathbf{x}_m)} \quad (3)$$

を用いる [Nadaraya 64][Watson 64]。関数 $k(\cdot, \cdot)$ は、動径基底関数 (RBF) であるカーネル関数とする。

式 (1) と $\mathbf{x}$ の事前分布 $p(\mathbf{x}|\boldsymbol{\lambda})$ を用いて、ノイズを除いた入力 $\mathbf{x}$ の事後分布は、ベイズの定理より

$$p(\mathbf{x}|\mathbf{u}) = \frac{p(\mathbf{u}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{u}|\mathbf{x})p(\mathbf{x})d\mathbf{x}} \quad (4)$$

となる。ここでは分布のパラメータを省略した。この事後分布は、ノイズを含む入力 $\mathbf{u}$ で条件付けられたもとのノイズを除いた入力 $\mathbf{x}$ の確率分布である。したがって、 $\mathbf{u}$ の値が与えられれば $\mathbf{x} \sim p(\mathbf{x}|\mathbf{u})$ となる。

ノイズあり入力 $\mathbf{u}$ が与えられたもとでは、ノイズなし入力 $\mathbf{x}$ は $p(\mathbf{x}|\mathbf{u})$ に従う確率変数であり、それを入力とする回帰関数 $\hat{g}(\mathbf{x})$ も確率変数となる。回帰関数 $\hat{g}(\mathbf{x})$ を確率変数変換の関数 $\hat{t} = \hat{g}(\mathbf{x})$ とすると、 $\hat{t}$ の分布は

$$p(\hat{t}|\mathbf{u}) = p(\hat{g}^{-1}(\hat{t})|\mathbf{u}) \left| \frac{d\hat{g}^{-1}(\hat{t})}{d\hat{t}} \right| \quad (5)$$

となる。式 (5) は、ノイズあり入力 $\mathbf{u}$ に対する回帰の予測分布となっている。これは分布の形で利用することもでき、その期待値

$$E[\hat{t}|\mathbf{u}] = \int \hat{t} p(\hat{t}|\mathbf{u}) d\hat{t} \quad (6)$$

は期待二乗損失最小の意味で最適な解となる。式 (6) が、本手法が与える回帰関数となる。

4.2 期待二乗損失最小

本提案手法が与える、ノイズを含む入力に対する回帰関数が、目的変数 t との期待二乗損失を最小にすることを示す。

目的変数 t と、ノイズを除いた入力 $\mathbf{x}$ に対する回帰関数 $y(\mathbf{x})$ との期待二乗損失を最小にする $y(\mathbf{x})$ は変分法により

$$y(\mathbf{x}) = \int t p(t|\mathbf{x}) dt \quad (7)$$

となる。Nadaraya-Watson カーネル回帰関数は、カーネル密度推定を利用している。カーネル関数の値が 0 よりも大きい値をもつ領域の体積を V とする。カーネル密度推定では、 V がデータ数 N の増加に従い徐々に 0 に近づくようにするなら、 $N \rightarrow \infty$ の極限で、推定された同時確率は真の同時確率に収束することが示せる [Scott 15]。したがって、そのような場合、Nadaraya-Watson カーネル回帰関数 $\hat{g}(\mathbf{x})$ は式 (7) と等しくなる。

目的変数 t に対し、ノイズを含む入力 $\mathbf{u}$ に対する回帰関数 $g(\mathbf{u})$ との損失を二乗損失 $L(t, g(\mathbf{u})) = (t - g(\mathbf{u}))^2$ とする。このとき、期待二乗損失 $E[L]$ は

$$E[L] = \int \int (t - g(\mathbf{u}))^2 p(t, \mathbf{u}) dt d\mathbf{u} \quad (8)$$

となり、変分法と式 (7) より、式 (8) を最小にする $g(\mathbf{u})$ は

$$g(\mathbf{u}) = \int \int t p(t|\mathbf{x}) p(\mathbf{x}|\mathbf{u}) dt d\mathbf{x} \quad (9)$$

$$= \int y(\mathbf{x}) p(\mathbf{x}|\mathbf{u}) d\mathbf{x} \quad (10)$$

となる。ノイズを除いた入力 $\mathbf{x}$ に対する回帰関数を $\hat{t} = y(\mathbf{x})$ とすると、式 (10) で表わされる $g(\mathbf{u})$ は

$$g(\mathbf{u}) = \int y(\mathbf{x}) p(\mathbf{x}|\mathbf{u}) d\mathbf{x} = E[\hat{t}|\mathbf{u}] = \int \hat{t} p(\hat{t}|\mathbf{u}) d\hat{t} \quad (11)$$

となる。式 (11) より、期待二乗損失最小の解である式 (10) と、ノイズを含む入力 $\mathbf{u}$ に対する回帰の予測分布の期待値である式 (6) が等しいことが分かる。

本手法では、Nadaraya-Watson カーネル回帰関数を基礎とし、目的変数の誤差を考慮して、目的変数とノイズを含む入力に対する回帰との期待二乗損失を最小にする回帰関数を与える。

*1 文献 [Bishop 95] では、Webb の手法について、入力をノイズあり入力ではなく、誤ってノイズなし入力としている。

4.3 時系列性をもつ説明変数とパラメータ推定

本提案手法を、説明変数が時系列性をもつ場合に適用する。訓練データ集合の説明変数値とノイズを除いた入力を HMM で表現し、入力ノイズ分布のパラメータを事後確率最大化により決定する。

本節では、現実のデータへの適用を考え、個人に対する健康診断と緊急検査の例で説明しよう。データとしては、血糖値とその血糖値に対する病気の危険度を想定する。血糖値と血糖値に対する危険度を想定する場合、体調がよい、悪いなどの患者個人の状態によって血糖の検査値は変動する。一般に、その状態には時系列性がある。本研究では、状態を HMM の潜在変数として表わす。そのため、 n を離散化された時間とし、状態を表現する潜在変数 z_n , $n = 1, \dots, N$, を導入する。訓練データ集合 D の説明変数値からなる集合を $\mathbf{X} = \{\mathbf{x}_n\}_{n=1, \dots, N}$ とし、潜在変数を用いて、 $\mathbf{X}$ と $\mathbf{x}_{N+1}$ を HMM で表現する。説明変数値の集合 $\mathbf{X}$ は健康診断時に得られた個人の血糖値を表わし、ノイズを除いた入力 $\mathbf{x}_{N+1}$ は緊急検査時の血糖値から食事や薬の影響などのノイズを取り除いた場合の値を表わす。ノイズを含む入力 $\mathbf{u} = \mathbf{x}_{N+1} + \boldsymbol{\eta}$ は緊急検査時の血糖値、ノイズ $\boldsymbol{\eta}$ は食事や薬などによる影響を表わす。また、目的変数値 t_n , $n = 1, \dots, N$, は、健康診断時に得られた個人の血糖値に対する病気の危険度を表わす。入力ノイズ分布のパラメータを $\boldsymbol{\kappa}$ とし、その真の値は与えられないとする。そのため、 $\boldsymbol{\kappa}$ を確率変数とし、 $\boldsymbol{\kappa} \sim p(\boldsymbol{\kappa}|\boldsymbol{\alpha})$ とする。分布 $p(\boldsymbol{\kappa}|\boldsymbol{\alpha})$ は $\boldsymbol{\kappa}$ の事前分布であり、 $\boldsymbol{\alpha}$ は超パラメータである。図 1 に、上記の変数間の関係を表わすグラフィカルモデルを示す。

最初に、超パラメータ $\boldsymbol{\alpha}$ の値を決定するために、 $\boldsymbol{\alpha}$ に関する周辺尤度を最大化する。その周辺尤度は

$$p(\mathbf{u}|\boldsymbol{\alpha}, \mathbf{X}) = \int \int p(\mathbf{u}|\mathbf{x}_{N+1}, \boldsymbol{\kappa})p(\boldsymbol{\kappa}|\boldsymbol{\alpha})p(\mathbf{x}_{N+1}|\mathbf{X})d\boldsymbol{\kappa}d\mathbf{x}_{N+1} \quad (12)$$

となる。ここで、 $p(\mathbf{x}_{N+1}|\mathbf{X})$ は、HMM から得られる $\mathbf{x}_{N+1}$ の予測分布である。式 (12) を最大にする $\boldsymbol{\alpha}$ を $\boldsymbol{\alpha}_{\text{MML}}$ とする。

次に、入力ノイズ分布のパラメータ $\boldsymbol{\kappa}$ の値を決定するために、 $\boldsymbol{\kappa}$ に関する事後確率を最大化する。パラメータ $\boldsymbol{\kappa}$ の事後分布は、 $\boldsymbol{\alpha}_{\text{MML}}$ を用いて

$$\begin{aligned} p(\boldsymbol{\kappa}|\mathbf{u}, \boldsymbol{\alpha}_{\text{MML}}, \mathbf{X}) &= \int \frac{p(\mathbf{u}|\boldsymbol{\kappa}, \mathbf{x}_{N+1})p(\boldsymbol{\kappa}|\boldsymbol{\alpha}_{\text{MML}})}{\int p(\mathbf{u}|\boldsymbol{\kappa}, \mathbf{x}_{N+1})p(\boldsymbol{\kappa}|\boldsymbol{\alpha}_{\text{MML}})d\boldsymbol{\kappa}} p(\mathbf{x}_{N+1}|\mathbf{X})d\mathbf{x}_{N+1} \\ &= \int p(\boldsymbol{\kappa}|\mathbf{u}, \mathbf{x}_{N+1}, \boldsymbol{\alpha}_{\text{MML}})p(\mathbf{x}_{N+1}|\mathbf{X})d\mathbf{x}_{N+1} \end{aligned} \quad (13)$$

となる。式 (13) を最大化して求めた $\boldsymbol{\kappa}$ を $\boldsymbol{\kappa}_{\text{MAP}}$ とする。

最後に、入力ノイズ分布のパラメータ $\boldsymbol{\kappa}_{\text{MAP}}$ を用いて、ベイズの定理より、ノイズを除いた入力 $\mathbf{x}_{N+1}$ の事後分布は

$$\begin{aligned} p(\mathbf{x}_{N+1}|\mathbf{u}, \boldsymbol{\kappa}_{\text{MAP}}, \mathbf{X}) &= \frac{p(\mathbf{u}|\mathbf{x}_{N+1}, \boldsymbol{\kappa}_{\text{MAP}})p(\mathbf{x}_{N+1}|\mathbf{X})}{\int p(\mathbf{u}|\mathbf{x}_{N+1}, \boldsymbol{\kappa}_{\text{MAP}})p(\mathbf{x}_{N+1}|\mathbf{X})d\mathbf{x}_{N+1}} \end{aligned} \quad (14)$$

となる。式 (14) により、訓練データ集合から構築される Nadaraya-Watson 回帰関数を確率変数化すると、その分布は $p(\hat{t}|\mathbf{u}, \boldsymbol{\kappa}_{\text{MAP}}, \mathbf{X})$ となり、その期待値をとると期待二乗損失最小の予測となる。

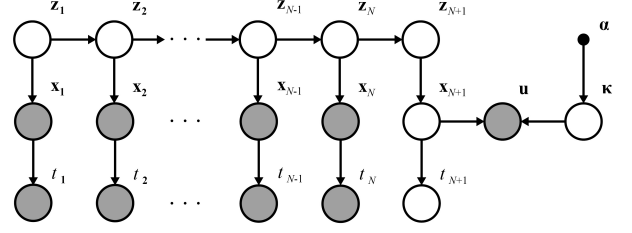


図 1: 説明変数が時系列性をもつ場合のグラフィカルモデル。

5. 評価実験

提案手法を評価するために、2 つの実験を行なう。実験では、回帰の予測に対する RMSE を評価指標とし、血糖値と血糖値に対する危険度を想定した人工データを用いる。その上で、提案手法と Webb の手法を比較する。

1 つ目の実験では、真の入力ノイズ分布が与えられるとする。ノイズなし入力 x に対する真の回帰関数は、線形関数 $f_1(x) = x/2$ とする。入力ノイズ分布は、 $p_\eta(\eta|\sigma) = \mathcal{N}(\eta|0.00, \sigma^2)$ とし、 $\sigma = 5.00, 20.00, 50.00$ の 3 つの値を用いる。仮定する x の事前分布は、 $\mathcal{N}(x|\mu, \lambda^{-1})$ とする。データ数 $N = 20$ とし、訓練データ集合の説明変数値を $\mathcal{N}(100.00, (50.00)^2)$ から得られるサンプル、訓練データ集合の目的変数値をその説明変数値に関数 $f_1(x)$ を適用し、 $\mathcal{N}(0.00, (20.00)^2)$ から得られる誤差を加えたものとする。テストデータは、訓練データ集合の説明変数値を生成した分布と同じ分布から得られる 10 個のサンプルに、入力ノイズ分布 $p_\eta(\eta|\sigma)$ から得られるノイズを加えたものとする。実験結果を図 2 に示す。パラメータ σ のすべての値で、Webb の手法よりも提案手法の精度がよいことが分かる。また、 $\sigma = 5.00$ の回帰関数の比較を図 3 に示す。Webb の手法の回帰関数よりも、提案手法の回帰関数では過学習が抑えられていることが分かる。

2 つ目の実験では、真の入力ノイズ分布は与えられないとし、説明変数は時系列性をもつとする。ノイズなし入力 x に対する真の回帰関数は、 $f_2(x) = 5\sqrt{2}x$ とする。この関数は、血糖値に対する危険度を参考に作成した危険度関数である [北村 14]。入力ノイズを生成する分布は、 $p(\eta|\sigma) = \mathcal{N}(\eta|0.00, \sigma^2)$ とし、 $\sigma = 5.00, 10.00, 20.00, 50.00$ の 4 つの値を用いる。データ数 $N = 20$ とし、訓練データ集合の説明変数値には、HMM から得られるサンプルを用いる。その HMM のパラメータは、状態数 $S = 2$ とし、初期状態確率 $\mathbf{P}$ 、状態遷移確率 $\mathbf{A}$ 、平均パラメータ $\mathbf{B}$ 、分散パラメータ $\mathbf{C}$ として、

$$\mathbf{P} = \begin{pmatrix} 0.40 & 0.60 \\ 0.60 & 0.30 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 0.60 & 0.40 \\ 0.30 & 0.70 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 50.00 \\ 150.00 \end{pmatrix}, \mathbf{C} = \begin{pmatrix} 200.00 \\ 500.00 \end{pmatrix}$$

を用いる。訓練データ集合の目的変数値は、その説明変数値に関数 $f_2(x)$ を適用し、 $\mathcal{N}(0.00, (10.00)^2)$ から得られる誤差を加えたものとする。テストデータは、訓練データ集合の説明変数値を生成した HMM と同じ HMM から得られる $N + 1$ 個目のサンプルに、入力ノイズ分布 $p_\eta(\eta|\sigma)$ から得られるノイズを加えたものとする。入力ノイズを生成する分布のパラメータ σ の各値ごとに 5 つのデータセットを作成し、5 回の実験結果の平均を Webb の手法と比較する。実験結果を図 4 に示す。パラメータ $\sigma = 5.00$ では、Webb の手法よりも提案手法の精度が悪いことが分かる。一方、それ以外の値では、Webb の手法よりも提案手法の精度がよいことが分かる。

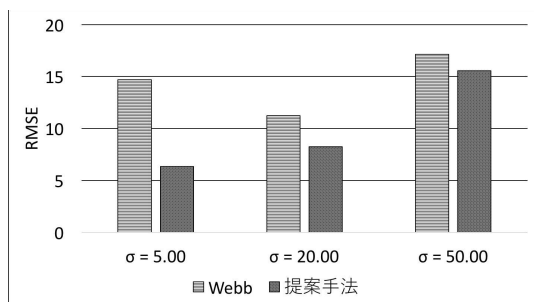


図 2: 提案手法と Webb 手法との結果の比較.

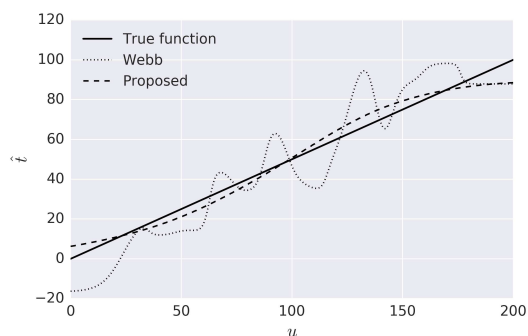


図 3: 提案手法と Webb 手法との回帰関数の比較.

6. 考察

真の入力ノイズ分布が与えられる場合、ノイズを生成した分布のパラメータ σ の値が小さいほど、すなわち、入力に含まれるノイズが小さいほど Webb の手法に比べて提案手法の精度がよいことが図 2 から分かる。Webb の回帰手法は、新たなノイズあり入力 u の値に対して、説明変数値との近さでその説明変数値に対応する目的変数値を重み付けて回帰を行なう。目的変数値には誤差が含まれているため、その誤差に追従してしまい、提案手法に比べて精度が下がったと考える。図 3 の $\sigma = 5.00$ で得られる回帰関数から、 σ の値が小さい場合、すなわち、ノイズが小さい場合はその影響が特に大きいと考える。パラメータ $\sigma = 20.00, 50.00$ では、Webb の手法でも過学習は抑えられる。その理由は、目的変数値に誤差が含まれていたとしても、広い範囲の多くの説明変数値に重みが与えられ、誤差への追従が平均化されるためだと考える。一方、提案手法では、どのような σ の値でも、目的変数値の誤差を考慮しているため、誤差への追従は抑えられ精度がよくなると考える。

真の入力ノイズ分布が与えられない場合、説明変数の時系列性を利用することにより、Webb の手法よりも提案手法の精度が全体的によくなることが図 4 の結果から分かる。入力ノイズを生成する分布のパラメータ $\sigma = 5.00$ の場合、Webb の手法よりも提案手法の精度が少し劣る。この理由について考える。入力 u に含まれるノイズが小さいため、 u に近い説明変数値の確率を高くすればよい。しかし、入力ノイズ分布のパラメータ θ の推定値は σ の値に比べて大きいことから、 x_{N+1} の事後分布が事前分布に近づき、 u に近い説明変数値の確率が低くなるため、精度が悪くなると考える。一方、Webb の手法では u から遠い説明変数値にも重みを与えるようになるが、 u に近い説明変数値の重みが強いいため、入力に含まれるノイズが小さい場合には、提案手法に比べて精度がよくなったと考える。

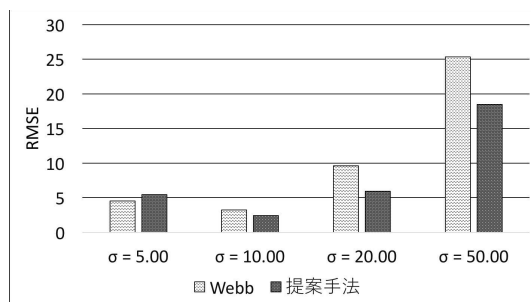


図 4: 説明変数が時系列性をもつ場合の提案手法と Webb 手法との結果の比較.

7. おわりに

本研究では、ノイズを含む入力に対する回帰手法を提案した。その手法は、カーネル密度推定を利用したノイズなし入力に対する回帰関数である Nadaraya-Watson カーネル回帰関数を基礎とする。その上で、目的変数の誤差を考慮し、目的変数とノイズを含む入力に対する回帰との期待二乗損失を最小にする回帰関数を与える。

今後の課題として、HMM の状態数を確率変数化し、中華料理店過程によって決定することが挙げられる。

参考文献

- [Bishop 95] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- [濱崎 08] 濱崎直孝・高木康 (2008). 臨床検査の正しい仕方－検体採取から測定まで－. 宇宙道八木書店.
- [Kamatani 10] Kamatani, Y., K. Matsuda, Y. Okada, M. Kubo, N. Hosono, Y. Daigo, Y. Nakamura, and N. Kamatani (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature Genetics*, 42, 210-215.
- [北村 14] 北村聖・中村丁次 (2014). からだの検査数値：70 項目から見える病気のサイン. 高久史磨監修，ニュートンプレス.
- [Nadaraya 64] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9(1), 141-142.
- [Scott 15] Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Oxford University Press.
- [下方 13] 下方浩史・安藤富士子 (2013). 検査基準値の考え方－医学における正常と異常－. *日本老年医学会雑誌*, 50(2), 168-171.
- [Watson 64] Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A* 26, 359-372.
- [Webb 94] Webb, A. R. (1994). Functional approximation by feed-forward networks: a least-squares approach to generalization. *IEEE Transactions on Neural Networks*, 5(3), 363-371.