P-8

ベイズ推定における回帰問題について
Regression problem in Bayesian estimation

○渡邉元気 [1], 青柳美輝 [2]
Genki Watanabe, Miki Aoyagi

Abstract: Regression analysis is one of the important typical tasks in image or speech recognition, economics and so on. In this paper, we consider the generalization error of regression problem in Bayesian estimation, which describes how precisely the predictive function approximates the true density function.

1. はじめに

近年，画像認識や音声認識，経済学などに応用されている機械学習の代表的なタスクの１つとして回帰手法がある．回帰とは入力から出力を予測するための関数をデータから求めるタスクである．ここでは，ベイズ推測における回帰問題について述べる．ベイズ推測においては，推測関数の誤差を測るための汎化誤差が重要である．汎化誤差を求めるためには真の分布が必要になるが，現実の問題では真の分布は不明である．本稿では回帰問題において，真の分布が未知の場合，サンプルから得られる経験誤差から汎化誤差を推測する手法について述べる．

2. 回帰関数

回帰の目的は，入力 x と出力 y からなる訓練データ集合が与えられたとき，新しい入力 X から対応する Y の値を予測することである．そのために，新しい入力 X から対応する Y の値の予測となるような関数 $f(x)$ のうち，誤差が最も小さくなるようなものを選ぶことを目指す．誤差を測る関数として，回帰でよく使われているのは二乗誤差である．二乗誤差を最小にするものとして，次で定義する回帰関数がある．

定義 1 (回帰関数) X, Y を確率変数とする．
$X = x$ のときの Y の平均値

$$E[Y|x] = \int yp(y|x)dy$$

を x から y への回帰関数という．

定理 2 関数 $y = f(x)$ が与えられたとき，その二乗誤差を表す汎関数を

$$E[(Y - f(X))^2] = \int\int(y - f(x))^2 p(x, y)dxdy$$

とすると，$E[(Y - f(X))^2]$ は $f(x) = E[Y|x]$ のとき最小になる．

(証明)

$$\begin{aligned}
&E[(Y - f(X))^2] \\
&= \int\int(y - f(X))^2 p(x, y)dxdy \\
&= \int\int\{y - E[Y|x] + E[Y|x] - f(x)\}^2 p(x, y)dxdy \\
&= \int\int\{\{y - E[Y|x]\}^2 + 2\{y - E[Y|x]\}\{E[Y|x] - f(x)\} \\
&\qquad + \{E[Y|x] - f(x)\}^2\}p(y|x)p(x)dxdy \\
&= \int\int\{y - E[Y|x]\}^2 p(y|x)dyp(x)dx \\
&\qquad + \int\{E[Y|x] - f(x)\}^2 p(x)dx
\end{aligned}$$

よって，$f(x) = E[Y|x]$ のとき最小になる．□

3. ベイズ推測における回帰関数の推定

確率変数 (X, Y) を $\mathbb{R}^M \times \mathbb{R}^N$ に値をとるものとし，その確率密度関数を

$$q(x, y) = \frac{q(x)}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp(-\frac{|y - r_0(x)|^2}{2\sigma^2})$$

とする．ここで $q(x)$ は $X \in \mathbb{R}^M$ の確率密度関数，$\sigma > 0$ は定数，$r_0(x)$ は $\mathbb{R}^M$ から $\mathbb{R}^N$ への関数とする．関数 $r_0(x)$ を真の回帰関数と呼ぶ．また，$|\cdot|$ は $\mathbb{R}^N$ におけるノルムを表す．$\{(X_i, Y_i); i = 1, 2, ..., n\}$ を $q(x, y)$ に従う独立なサンプルとする．パラメータの集合を $W \subset \mathbb{R}^d$ とし，$r(x, w)$ を $\mathbb{R}^M \times W$ から $\mathbb{R}^N$ への関数とする．$H(w)$ を

$$H(w) = \frac{1}{2}\sum_{i=1}^{n}|Y_i - r(X_i, w)|^2$$

とする．パラメータ w の関数 $F(w)$ が与えられたとき，平均をとる操作 $E_w[\ ]$ を

$$E_w[F(w)] = \frac{\int F(w)\exp(-\beta H(w))\phi(w)dw}{\int \exp(-\beta H(w))\phi(w)dw}$$

と表記する．$\phi(w)$ は w の事前分布，β は逆温度と呼ばれる定数である．$E_w[\ ]$ を用いて計算される回帰関数

$$y = E_w[r(x, y)]$$

を回帰関数のベイズ推測と呼ぶことにする．

1: 日大理工・院 (前)・数学　2: 日大理工・教員・数学

定義 3 (汎化誤差, 経験誤差)

$$G = \frac{1}{2} E_X E_Y[|Y - E_w[r(X,w)]|^2]$$

$$T = \frac{1}{2n} \sum_{i=1}^{n} |Y_i - E_w[r(X_i,w)]|^2$$

をそれぞれ，汎化誤差，経験誤差という．

もし，フィッシャー情報行列

$$I_{ij}(w) = \int \frac{\partial}{\partial w_i} r(x,w) \cdot \frac{\partial}{\partial w_j} r(x,w) q(x) dx$$

が任意の $w \in W$ で正定値であれば正則回帰問題といい，そうでないとき，すなわち，ある $w \in W$ が存在して，$\det I(w) = 0$ であるとき，特異回帰問題といわれる．次の定理 6 と定理 7 は正則でも特異でも成り立つ．

定義 4 $K(w) = \frac{1}{2} \int |r(x,w) - r_0(x)|^2 q(x) dx$ と定義する．

$K(w)$ が解析関数であれば

$$\zeta(z) = \int_W K(w)^z \varphi(w) dw$$

は Re$(z) > 0$ において正則関数であり，有理型関数として複素平面全体に解析接続することができる．その極はすべて負の有理数である．$\zeta(z)$ の最大の極を $(-\lambda)$ とし，位数を m とする．
定数 $\lambda > 0$ を実対数閾値と呼ぶ．

定義 5

$$V_n = \sum_{i=1}^{n} (E_w[|r(X_i,w)|^2] - |E_w[r(X_i,w)]|^2)$$

とおくと，ある定数 $\nu = \nu(\beta) > 0$ が存在して

$$\lim_{n \to \infty} E[V_n] = \frac{2\nu}{\beta}$$

が成り立つ．定数 ν を特異ゆらぎと呼ぶ．

定理 6 $S = \frac{N\sigma^2}{2}$ とする．ある $w_0 \in W$ が存在して $r(x,w_0) = r_0(w)$ のとき，

$$\lim_{n \to \infty} n(E[G] - S) = \frac{\lambda - \nu}{\beta} + \nu\sigma^2$$

$$\lim_{n \to \infty} n(E[T] - S) = \frac{\lambda - \nu}{\beta} - \nu\sigma^2$$

が成り立つ．

正則回帰問題の場合は，

$$\lim_{n \to \infty} n(E[G] - S) = \frac{d\sigma^2}{2}$$

$$\lim_{n \to \infty} n(E[T] - S) = -\frac{d\sigma^2}{2}$$

となる．

実対数閾値 λ と 特異ゆらぎ ν が特異回帰問題のベイズ推測において重要な値であることが分かる．

定理 7

$$E[G] = E[(1 + \frac{2\beta V_n}{nN})T] + o_n$$

ここで，o_n は n の関数であり，$no_n \to 0$ $(n \to \infty)$ を満たす．

補題 8 $E[V_n T] - E[V_n]E[T] \to 0$ $(n \to \infty)$.

(定理 7 の証明)

定理 6 より，

$$E[G] = \frac{N\sigma^2}{2} + (\frac{\lambda - \nu}{\beta} + \nu\sigma^2)\frac{1}{n} + o_n$$

$$E[T] = \frac{N\sigma^2}{2} + (\frac{\lambda - \nu}{\beta} - \nu\sigma^2)\frac{1}{n} + o_n.$$

よって

$$E[G] - E[T] = \frac{2\nu\sigma^2}{n} + o_n$$

$$E[G] = E[T] + \frac{2\nu\sigma^2}{n} + o_n$$

$$= E[T](1 + \frac{2\beta E[V_n]}{Nn}) + o_n$$

であることと，補題 8 より定理を得る．□

V と T は真の分布が未知の場合でも，サンプルから計算できるので，定理 5 によりサンプルから $E[G]$ が得られる．

4. 参考文献

[1] S. Watanabe:“ A Limit Theorem in Singular Regression Problem ”,Advanced Studies in Pure Mathematics 57,473-492 (2010)

[2] 渡辺澄夫:“ベイズ統計の理論と方法”, コロナ社 (2012)

[3] ビショップ,C.M.:“ パターン認識と機械学習 (上) ”, シュプリンガー (2007)