データ解析レポート課題・第一

14_01043　伊澤　侑祐

問 1　計算問題

(1)　$E(a)$ を最小にするパラメータ $a^* = \{(a^*)_k\}$ を求めよ

まず、k を固定する。二乗誤差

$$
\begin{aligned}
E(a) &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - f(x_i, a))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \sum_{k=1}^{K} a_k e_k(x))^2
\end{aligned}
\tag{1}
$$

を a_k について偏微分し、$\frac{\partial E}{\partial a_k} = 0$ を解く。

$$
\frac{\partial E}{\partial a_k} = \frac{2}{n}\sum_{i=0}^{n}(Y_i - a_k e_k(x)) \cdot e_k(x) = 0 \tag{2}
$$

$$
\therefore \sum_{i}^{n} Y_i e_k(x) - n a_k = 0 \tag{3}
$$

よって求める答えは式 3より

$$
a^* = \frac{1}{n}\sum_{i}^{n} Y_i e_k(x) \tag{4}
$$

である。

(2)　a^* の平均を求めよ

平均を $E_A(\cdot)$ で表す。$E_A(Y_i) = 0$ を用いて、

$$
\begin{aligned}
E_A(a^*) &= E_A\left(\frac{1}{n}\sum_{i=0}^{n} Y_i e_k(x)\right) \\
&= 0
\end{aligned}
\tag{5}
$$

となる。

(3)

まず、一般に (k, l) 成分の場合の共分散を考える。i 成分の期待値を μ_i と置くと、

$$
\begin{aligned}
\sigma_{i,j} &= E_A\left((a_k^* - \mu_i)(a_l^* - \mu_l)\right) \\
&= E_A\left(a_k^* \cdot a_l^*\right) \\
&= E_A\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i e_k(x_i)\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i e_l(x_i)\right)\right) \\
&= \frac{1}{n^2} E_A\left(\sum_{i,j=1}^{n} Y_i Y_j e_k(x_i) e_l(x_j)\right)
\end{aligned}
\tag{6}
$$

となる。ここで、$E_A(Y_i) = 0, E_A(Y_i^2) = 1$ であることより、

$$
\begin{aligned}
E_A\left(\sum_{i,j=1}^{n} Y_i Y_j e_k(x_i) e_l(x_j)\right) &= \sum_{i,j=1}^{n} E_A(Y_i) E_A(Y_j) \sum_{i,j=1}^{n} e_k(x_i) e_l(x_j) \\
&= \begin{cases} \sum_{i=1}^{n} E_A(Y_i^2) \sum_{i=1}^{n} e_k(x_i) e_l(x_i) & (i = j) \\ 0 & (i \neq j) \end{cases} \\
&= \begin{cases} n^2 & (i = j) \\ 0 & (i \neq j) \end{cases}
\end{aligned}
\tag{7}
$$

となる。したがって、求める分散共分散行列 Σ は、(6) と (7) より

$$
\Sigma = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = I \tag{8}
$$

と求まる。

(4)

問 2　応用問題

(1)

「市町村 2012estat.csv」に対し、回帰分析、主成分分析とクラスタ分析を用いて解析を行った。

(1.1)　回帰分析