

# 2023 年度 データサイエンス基礎 第 8 回課題

理工学研究科物質科学専攻物理学 PG

23MS113 石田洋音

1 日の電力需要量を目的変数として、説明変数には以下の 10 種類を考慮し、次のようなモデルを立てる。重回帰によりパラメータ  $\alpha_0, \alpha_1, \dots, \alpha_{10}$  を予測する。

$$z = \alpha_0 + \sum_{i=1}^{10} \alpha_i x_i \quad (1)$$

$x_1$ : 平均気温  
 $x_2$ : 最高気温  
 $x_3$ : 最低気温  
 $x_4$ : 降水量  
 $x_5$ : 日照時間  
 $x_6$ : 降雪量  
 $x_7$ : 平均風速  
 $x_8$ : 平均蒸気圧  
 $x_9$ : 平均湿度  
 $x_{10}$ : 平均現地気圧

それぞれの説明変数を「使う」・「使わない」で合計  $2^{10} - 1 = 1023$  通り (全て選ばない 1 通りは除く) のモデルを作り適切な説明変数の組みを探した。

## 交差検証法

パラメータの推定には 2 日から 30 日までのデータを用いて、1 日と 31 日のデータをテストに利用する交差検証を行う。1 日と 31 日のデータでモデルによる予測値と実際の値の相対誤差を求め、相対誤差の平均をそのモデルの評価値とする。

各モデルの評価値は次のようになる。

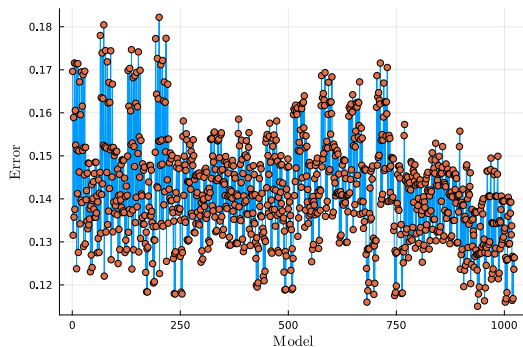


図 1 評価値

おおよそ相対誤差が 11 % から 19 % 程度発生していることがわかる。この中で最も評価値の小さいモデルは

- ・ 平均気温
- ・ 最高気温

- ・ 最低気温
- ・ 日照時間
- ・ 平均風速
- ・ 平均湿度

の 6 つを採用したもので、評価値は 11.5 % である。説明変数のパラメータは

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \\ \alpha_0 \end{pmatrix} = \begin{pmatrix} 1844 \\ -724 \\ -1557 \\ 0 \\ -1149 \\ 0 \\ 2394 \\ 0 \\ 225 \\ 0 \\ 70660 \end{pmatrix} \quad (2)$$

## 情報基準量 (AIC)

次に情報基準量 (AIC) の観点から最適な説明変数の組を考える。AIC を次で定義する。  $n$  はサンプルサイズ (今回は 31),  $S_e$  は残差二乗和,  $p$  は説明変数の数である。

$$AIC = n \left( \log \left( 2\pi \frac{S_e}{n} \right) \right) + 2(p + 2) \quad (3)$$

モデルごとの AIC の値は次のようになる。

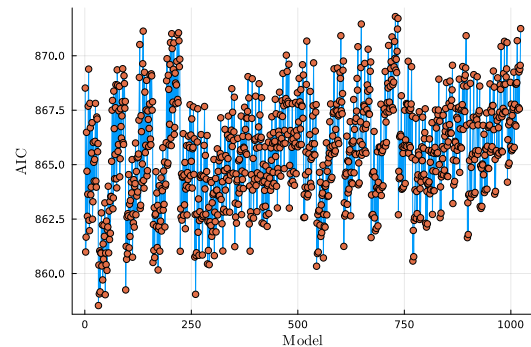


図 2 AIC

最も AIC の小さいモデルは、説明変数として日照時間のみを用いるもので、パラメータは

$$\begin{aligned} \alpha_0 &= 92145 \\ \alpha_5 &= -1272 \end{aligned}$$

であった。