

Stepwise (ステップワイズ) 法による

説明変数 (入力変数・記述子・

特徴量) の選択

明治大学 理工学部 応用化学科

データ化学工学研究室 金子 弘昌

Stepwise (ステップワイズ) 法とは？

- ✓ 説明変数 (入力変数・記述子・特徴量) を選択する手法
- ✓ 1つずつ説明変数を追加したり、削除したりしながら、最適な説明変数の組合せを探す
- ✓ 回帰モデルの構築を繰り返す数が多くなると時間がかかる
- ✓ どんな回帰分析手法とも組み合わせることができる

Stepwise法の種類

- ✓変数増加法 (forward stepwise)
 - 説明変数なしからはじめて、1 つずつ説明変数を増やす

- ✓変数減少法 (backward stepwise)
 - すべての説明変数からはじめて、1 つずつ説明変数を減らす

- ✓変数増減法 (forward-backward stepwise)
 - 説明変数なしからはじめて、1 つずつ説明変数を増やすか減らすかする

- ✓変数減増法 (backward-forward stepwise)
 - すべての説明変数からはじめて、1 つずつ説明変数を増やすか減らすかする

どのように説明変数を増やすか？

- ✓ある説明変数の組み合わせが選ばれているとき
(最初はもちろん説明変数なし)
- ✓選ばれていない説明変数の中から、1つ選んで追加して、**評価指標**の値を計算する
- ✓計算し終わったら、追加した説明変数を戻す
- ✓選ばれていない説明変数すべてで**評価指標**の計算を行い、**評価指標**の値が一番もっとも良くなった説明変数を実際に追加する

評価指標 (最小二乗法による重回帰分析用) 1/2 ⁴

- ✓ Mallows's C_p
- 小さいほど良い

$$C_p = \frac{SSE}{S^2} - n + 2m$$

$$SSE = \sum_{i=1}^n \left(y^{(i)} - y_{\text{EST}}^{(i)} \right)^2$$

n : サンプル数

m : 回帰モデルを構築した
説明変数の数

S^2 : すべての説明変数を用いて
回帰分析を行ったときの
誤差の二乗の平均

$y^{(i)}$: i 番目のサンプルにおける
目的変数の値

$y_{\text{EST}}^{(i)}$: i 番目のサンプルにおける
目的変数の推定値

評価指標 (最小二乗法による重回帰分析用) 2/2 ⁵

✓赤池情報量規準 (Akaike's Information Criterion, AIC)

- 小さいほど良い

$$AIC = m \log \left(\frac{SSE}{m} \right) + 2$$

✓Bayesian Information Criterion (BIC)

- 小さいほど良い

$$BIC = m \log \left(\frac{SSE}{m} \right) + n \log(m)$$

評価指標（任意の回帰分析手法で使える）

- ✓ $RMSE_{CV}$: クロスバリデーション後のRoot Mean Squared Error
- 小さいほど良い

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^n \left(y^{(i)} - y_{CV\text{EST}}^{(i)} \right)^2}{n}}$$

$y_{CV\text{EST}}^{(i)}$: i 番目のサンプルにおける
クロスバリデーション後の
目的変数の推定値

- ✓ MAE_{CV} : クロスバリデーション後のMean Absolute Error (MAE)
- 小さいほど良い

$$MAE_{CV} = \frac{\sum_{i=1}^n \left| y^{(i)} - y_{CV\text{EST}}^{(i)} \right|}{n}$$

どのように説明変数を減らすか？

- ✓ある説明変数の組み合わせが選ばれているとき
(最初はもちろん説明変数なし)
- ✓選ばれている説明変数の中から、1つ選んで削除して、**評価指標**の値を計算する
- ✓計算し終わったら、削除した説明変数を戻す
- ✓選ばれている説明変数すべてで**評価指標**の計算を行い、**評価指標**の値が一番もっとも良くなった説明変数を実際に追加する

減らすときだけで使える手法

- ✓ 回帰分析手法の標準回帰係数に基づく変数削除
 - 標準回帰係数：標準化（オートスケーリング）後に計算された回帰係数
 - 選ばれている説明変数を用いて一度線形回帰分析を行い、標準回帰係数の絶対値がもっとも小さい変数を削除
 - 説明変数を1つ減らすときに、削除して回帰モデル構築を繰り返さなくてよいため、計算時間が短い

どのように説明変数を増やすか減らすか？⁹

- ✓ある説明変数の組み合わせが選ばれているとき
(最初はもちろん説明変数なし)
- ✓選ばれていない説明変数の中から、1つ選んで追加して、**評価指標**の値を計算する
- ✓計算し終わったら、追加した説明変数を戻す
- ✓選ばれている説明変数の中から、1つ選んで削除して、**評価指標**の値を計算する
- ✓計算し終わったら、削除した説明変数を戻す
- ✓全通りで**評価指標**の計算を行い、**評価指標**の値が一番もっとも良くなった説明変数を実際に追加 or 削除する