リッジ回帰(Ridge Regression, RR)

Least Absolute Shrinkage and Selection Operator (LASSO)

Elastic Net (EN)

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

RR・LASSO・EN とは？

✓線形の回帰分析手法

✓目的変数の誤差の二乗和に加えて、それぞれ以下の項を最小化することで、過学習を防ぐ

✓RR: 回帰係数の二乗和

✓LASSO: 回帰係数の絶対値の和

✓EN: 回帰係数の二乗和と絶対値の和 (RRとLASSOとの中間)

✓LASSOとENは回帰係数の値が0になりやすく、変数選択としても利用できる

OLS・RR・LASSO・EN・SVR

✓最小二乗法による線形重回帰分析
(Ordinary Least Squares, OLS)

✓リッジ回帰 (Ridge Regression, RR)

✓Least Absolute Shrinkage and Selection Operator (LASSO)

✓Elastic Net (EN)

✓サポートベクター回帰
(Support Vector Regression, SVR)

OLS・RR・LASSO・EN・SVRの共通点

✓線形の回帰分析手法

- たとえば説明変数が 2 つのとき、目的変数・説明変数をオートスケーリングしたあと、

$$y = x_1 b_1 + x_2 b_2 + f$$
$$= y_C + f$$
$$\left(y_C = x_1 b_1 + x_2 b_2 \right)$$

と表わされる

y： 目的変数

x_1, x_2： 説明変数 (記述子)

b_1, b_2： (標準)回帰係数

y_C： y の、x で表すことができる部分

f： y の、x で表すことができない部分 (誤差、残差)

✓ある関数 G を最小化することで回帰係数を求める

OLS・RR・LASSO・EN・SVRの違い 1/2

✓OLS: G は誤差の二乗和

$$G = \sum_{i=1}^{n} f_i^2 = \left\| \mathbf{y} - \mathbf{X}\mathbf{b} \right\|^2$$

n : サンプル数

f_i : i 番目のサンプルの誤差

行列の表し方については[こちら]

✓RR: G は誤差の二乗和と回帰係数の二乗和

$$G = \left\| \mathbf{y} - \mathbf{X}\mathbf{b} \right\|^2 + \lambda \sum_{i=1}^{m} b_i^2$$

m : 説明変数の数

b_i : i 番目の説明変数の回帰係数

λ : 重み

✓LASSO: G は誤差の二乗和と回帰係数の絶対値の和

$$G = \left\| \mathbf{y} - \mathbf{X}\mathbf{b} \right\|^2 + \lambda \sum_{i=1}^{m} \left| b_i \right|$$

OLS・RR・LASSO・EN・SVRの違い 2/2

✓EN: G は誤差の二乗和と回帰係数の二乗和と絶対値の和

$$G = \left\| \mathbf{y} - \mathbf{Xb} \right\|^2 + \lambda \left(\alpha \sum_{i=1}^{m} b_i^2 + (1-\alpha) \sum_{i=1}^{m} \left| b_i \right| \right)$$

α : 重み
($\alpha=1 \rightarrow$ RR,
$\alpha=0 \rightarrow$ LASSO)

✓SVR: G はある誤差関数 *h* と回帰係数の二乗和

$$G = h(\mathbf{y} - \mathbf{Xb}) + \lambda \sum_{i=1}^{m} b_i^2$$

- *h* についてはSVRの資料のときに

回帰係数の求め方

G が最小値を取る

⇩

G が極小値を取る

⇩

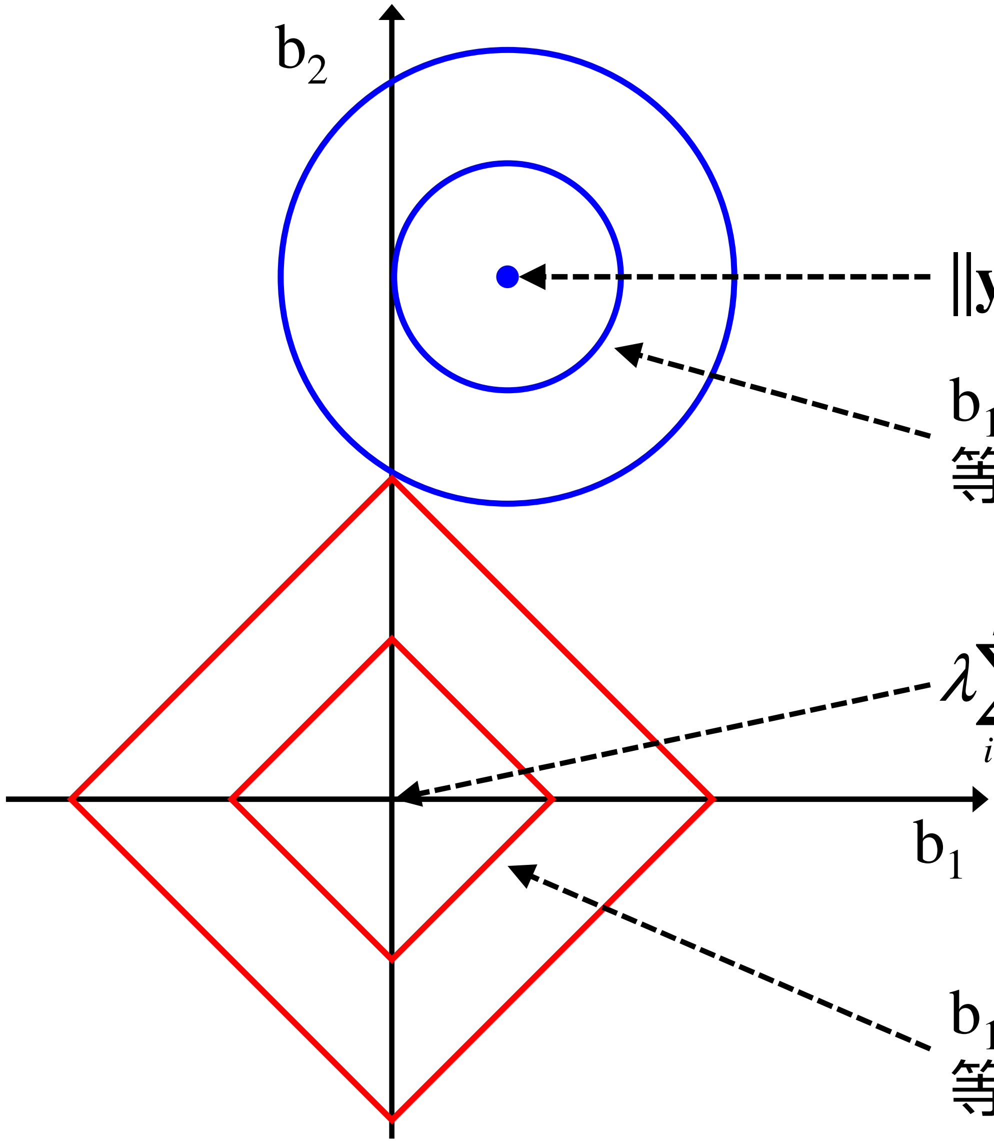
G を 各b_i で偏微分したものが 0　　$\dfrac{\partial G}{\partial b_i} = 0$

⇩

必要に応じて繰り返し計算により、 $\dfrac{\partial G}{\partial b_i} = 0$ を満たす各b_i を求める

どうしてLASSOは回帰係数が0になりやすいの？

$$G = \|\mathbf{y}-\mathbf{Xb}\|^2 + \lambda\sum_{i=1}^{m}|b_i|$$



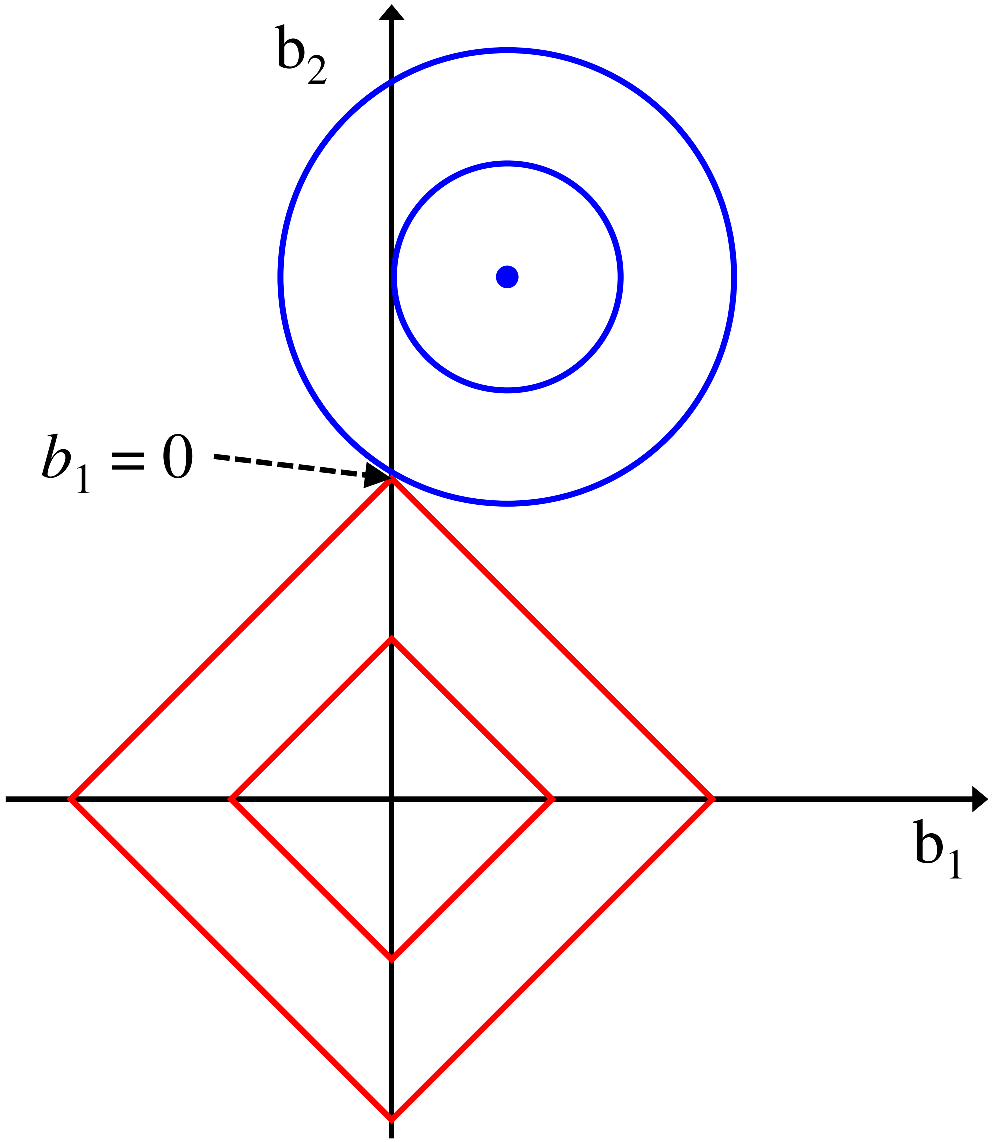
$\|\mathbf{y}\text{-}\mathbf{Xb}\|^2$ が**最小**になる (b_1, b_2)

b_1, b_2 を変えたときの$\|\mathbf{y}\text{-}\mathbf{Xb}\|^2$ の等高線

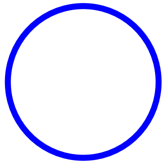
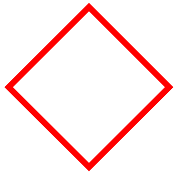
$\lambda\sum_{i=1}^{m}|b_i|$ が**最小**になる $(b_1, b_2) = (0,0)$

b_1, b_2 を変えたときの $\lambda\sum_{i=1}^{m}|b_i|$ の等高線

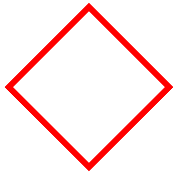
どうしてLASSOは回帰係数が0になりやすいの？[8]



$$G = \|\mathbf{y} - \mathbf{Xb}\|^2 + \lambda \sum_{i=1}^{m} |b_i|$$

 と  との交点が、

G が最小になる (b_1, b_2)

 の角が軸上にあるため

b_1 もしくは b_2 が 0 になりやすい

(ENも回帰係数が0になりやすい)

重み λ, α の決め方

✓グリッドサーチによって、クロスバリデーションの後の r^2 の値が
もっとも高い λ (RR, LASSO) もしくは λとαの組み合わせ (EN) とする

✓RRにおける λ の候補の例: $2^{-15}, 2^{-14}, \ldots, 2^{8}, 2^{9}$

✓LASSOにおける λ の候補の例: $2^{-15}, 2^{-14}, \ldots, 2^{-2}, 2^{-1}$

✓ENにおける λ の候補の例: $2^{-15}, 2^{-14}, \ldots, 2^{-2}, 2^{-1}$

✓ENにおける α の候補の例: 0, 0.01, …, 0.99, 1