正則化最小二乗法
(Ridge, lasso）

Regularized Least Squares Method

大阪府立大学 工学研究科
清水　悠生

はじめに

- ✓ 本記事の図の作成に使用したPythonコードは全てGitHubで公開しています

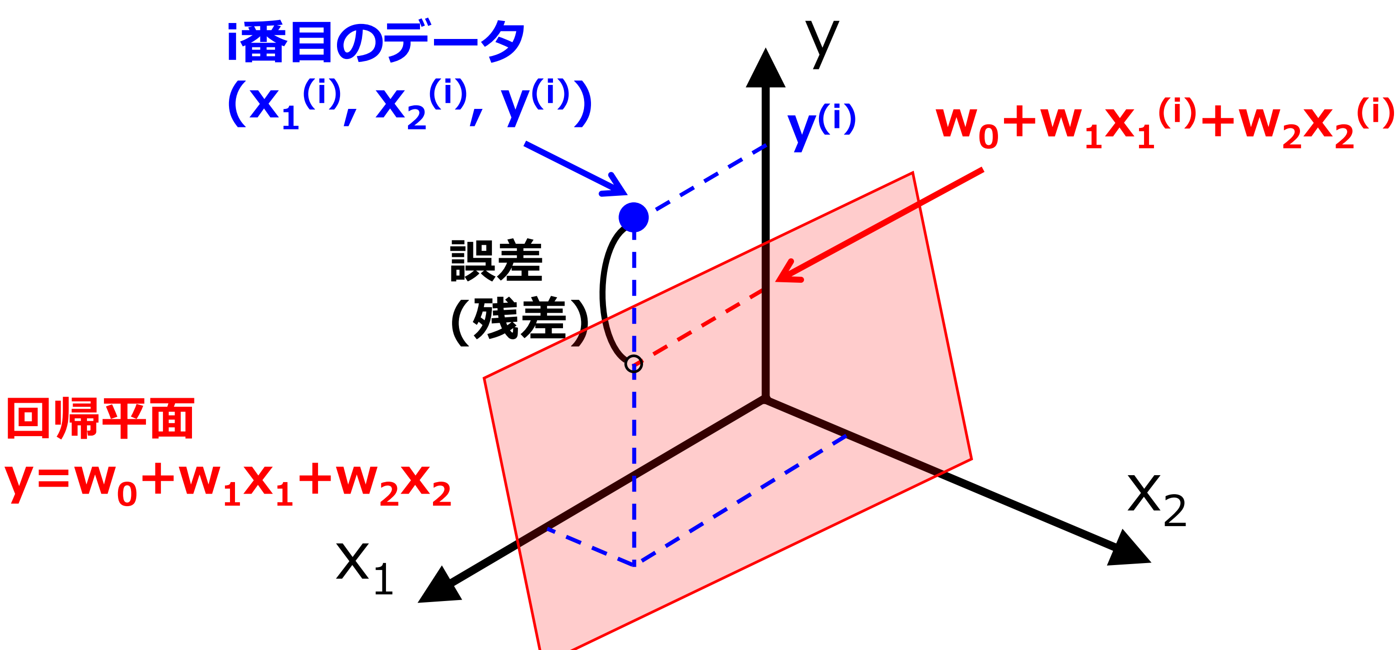
(復習)最小二乗法で扱う誤差関数

- ✓ 誤差関数を誤差の2乗の和とし
  誤差関数が最小となるような係数を計算する

(入力変数が2つの場合の)最小二乗法で扱う誤差関数E(w)

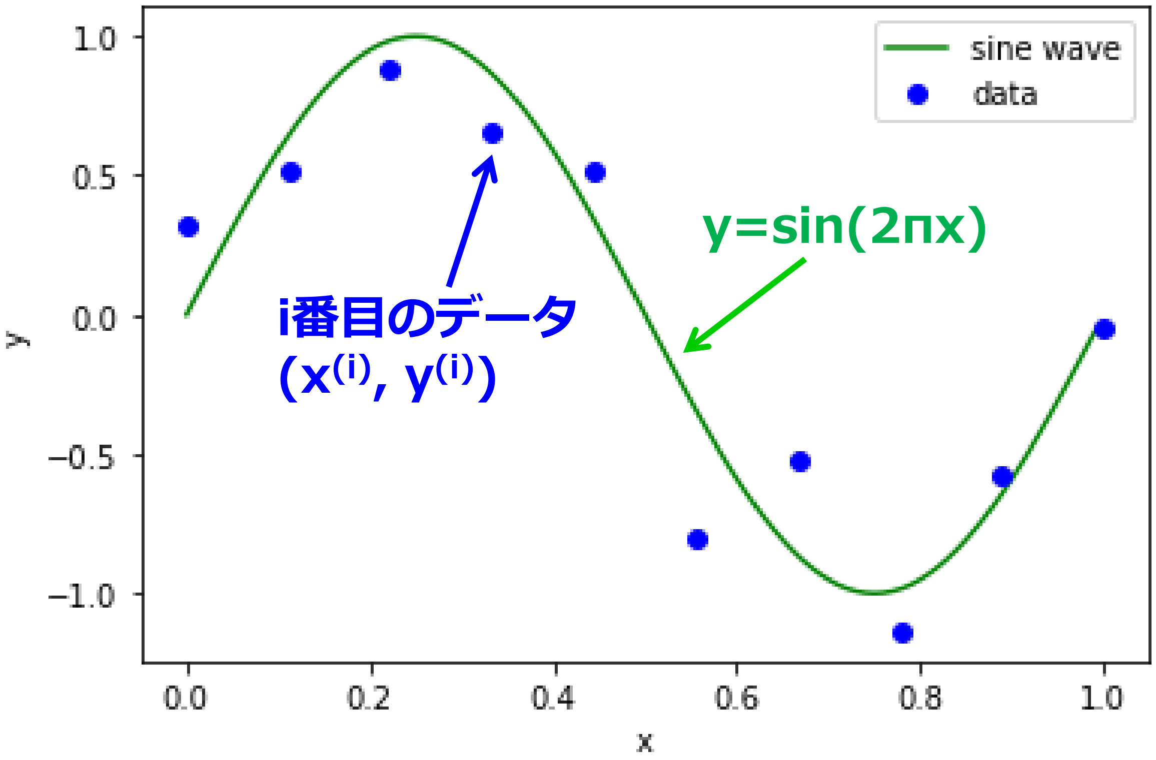
$$E(\boldsymbol{w}) = \sum_{i=1}^{n} \underbrace{\frac{1}{2}\left(y^{(i)} - \left(w_0 + w_1 x_1^{(i)} + w_2 x_2^{(i)}\right)\right)^2}_{\text{誤差の2乗}} \qquad \boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

n個のデータを仮定



多項式回帰の例

✓ サインカーブに正規分布ノイズを加えた10個のデータから多項式回帰分析を行う場合を考える



$x^{(i)}$は区間[0,1]で等間隔に10点生成

$y^{(i)}$は平均0, 分散0.04の正規分布ノイズを加味して，次式に従って生成

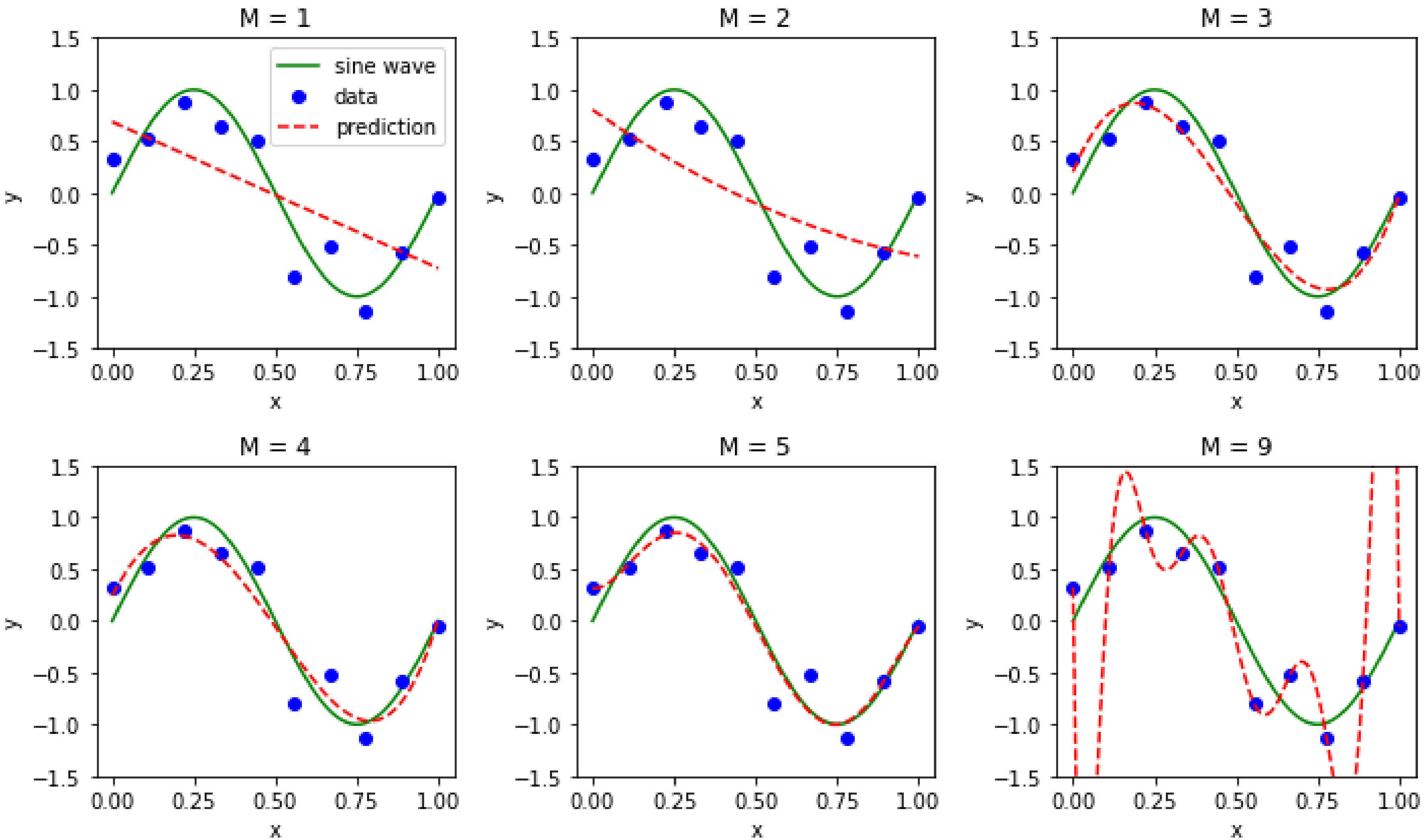
$$y^{(i)} = \sin(2\pi x^{(i)}) + N(0,0.04)$$

多項式回帰の例

多項式回帰式

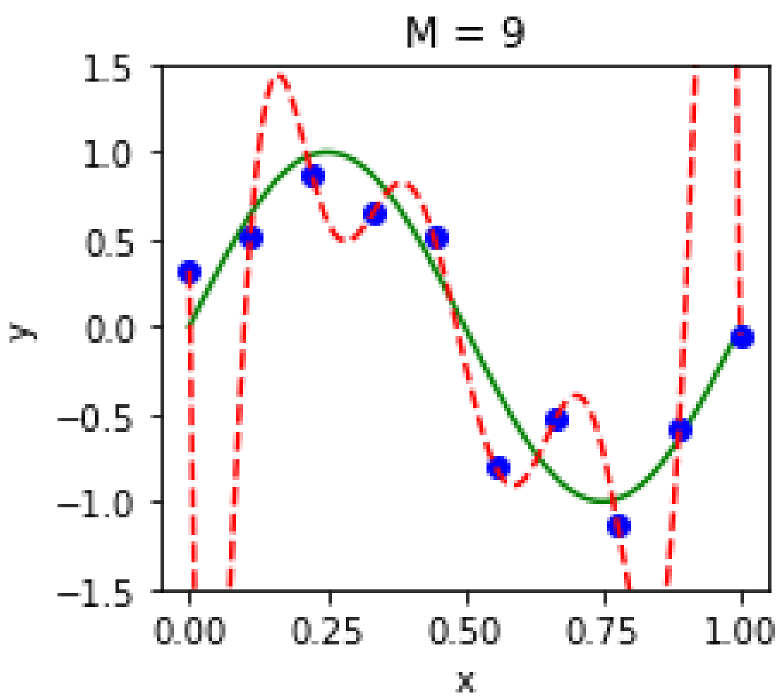
$$y = \sum_{i=0}^{M} w_i x^i$$

- ✓ 最小二乗法により多項式回帰の係数を計算
- ✓ M=3~5あたりでサインカーブへのあてはまりが良い



過学習とは

- ✓ M=9の例では回帰曲線が全てのデータを通っており
誤差関数E(**w***)=0となっている
- ✓ これは与えられたデータ数が10組なのに対し
学習する係数も10個存在するため
- ✓ 回帰曲線はノイズに過度に引きずられており
適切とは言えない⇒**過学習**と呼ばれる現象



$$y = w_0 + w_1 x^1 + w_2 x^2 + \cdots + w_8 x^8 + w_9 x^9$$

**与えられたノイズを含む点列に
過度に適合している
⇒過学習（過適合; over-fitting）**

過学習の際の係数の値

- ✓ 次数が大きくなるにつれ係数の値も大きくなる
- ✓ 直感的には，学習の自由度が増え
  ノイズの影響を受けやすくなるため

次数が増加すると
係数が大きくなる

	M=1	M=2	M=3	M=4	M=5	M=9
w0	0.69	0.80	0.21	0.24	0.32	0.32
w1	-1.42	-2.19	7.66	6.28	-0.47	-254.05
w2	nan	0.78	-25.19	-18.21	38.05	5954.13
w3	nan	nan	17.31	6.12	-152.90	-54168.85
w4	nan	nan	nan	5.60	187.87	257772.53
w5	nan	nan	nan	nan	-72.91	-712418.06
w6	nan	nan	nan	nan	nan	1183107.13
w7	nan	nan	nan	nan	nan	-1162608.56
w8	nan	nan	nan	nan	nan	622374.75
w9	nan	nan	nan	nan	nan	-139759.40

最小二乗法により得られた各回帰式の係数

正則化項の導入

- ✓ 誤差関数に**正則化項**を導入
- ✓ 誤差最小化時に係数も最小化されるので過学習を抑制できる
- ✓ 正則化項の次数q=2の時の回帰手法を**ridge回帰**
  q=1の時の回帰手法を**lasso**と呼ぶ

正則化最小二乗法で扱う誤差関数

$$\tilde{E}(\boldsymbol{w}) = \underbrace{E(\boldsymbol{w})}_{\text{二乗和誤差項}} + \underbrace{\frac{\alpha}{2}\sum_{i=0}^{M}|w_i|^q}_{\text{正則化項}}$$

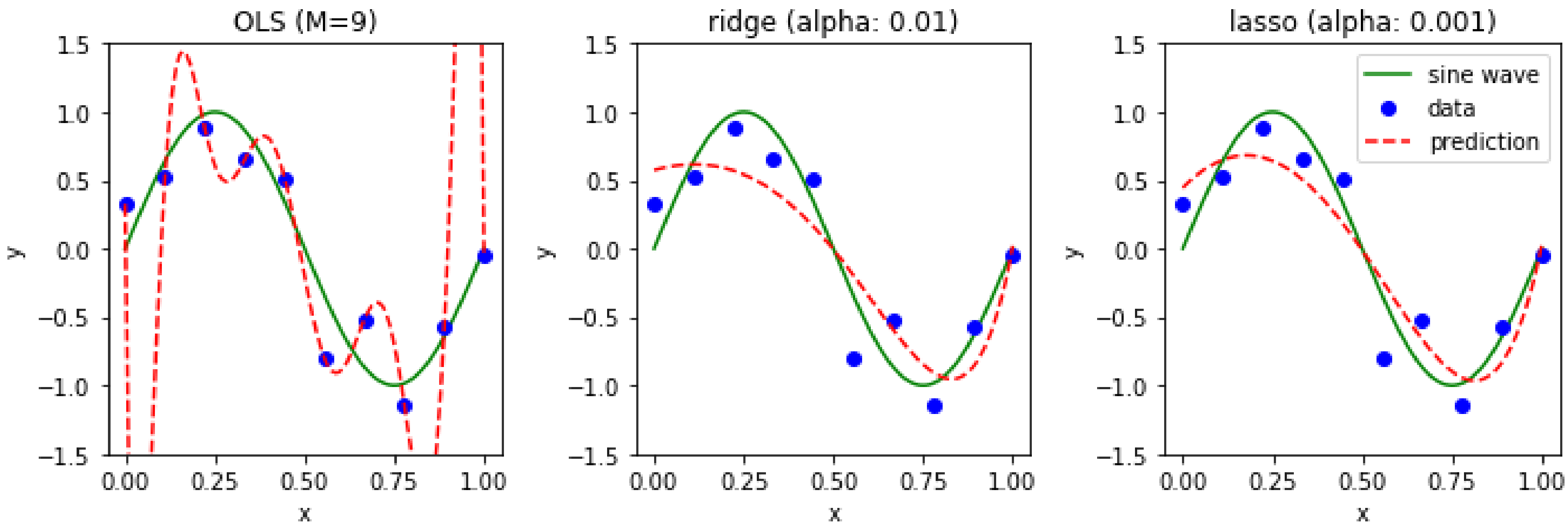
α: 二乗和誤差項と正則化項の相対的な重要度を調整するハイパーパラメータ
(パラメータ**w**を決定するためのパラメータなのでハイパーパラメータと呼ぶ)

正則化項の代表的な例

$$\sum_{i=0}^{M}|w_i|^q = \begin{cases} {w_0}^2 + {w_1}^2 + \cdots + {w_M}^2 & \textbf{q=2} \Rightarrow \textbf{ridge回帰} \\ |w_0| + |w_1| + \cdots + |w_M| & \textbf{q=1} \Rightarrow \textbf{lasso} \end{cases}$$

正則化項を導入した際の回帰曲線

- ✓ 正則化項を導入すると発振が抑えられ
サインカーブへのあてはまりが良くなることがわかる



正則化項を導入した9次多項式回帰の学習結果

正則化項を導入した際の回帰係数

- ✓ 正則化項を導入すると係数の値が小さく抑えられる
- ✓ **lassoにはいくつかの係数が0となる**
スパース（疎）な解が得られやすいという特徴がある

	OLS (M=9)	ridge	lasso
w0	0.3249	0.5736	0.4444
w1	-254.0475	0.7424	2.6888
w2	5954.1292	-3.0046	-7.6257
w3	-54168.8450	-1.8439	-0.0000
w4	257772.5254	-0.3024	0.0000
w5	-712418.0575	0.6704	2.1443
w6	1183107.1250	1.0621	2.3690
w7	-1162608.5621	1.0392	0.0000
w8	622374.7536	0.7569	0.0000
w9	-139759.3957	0.3271	-0.0000

lassoでは
スパースな解が
得られる

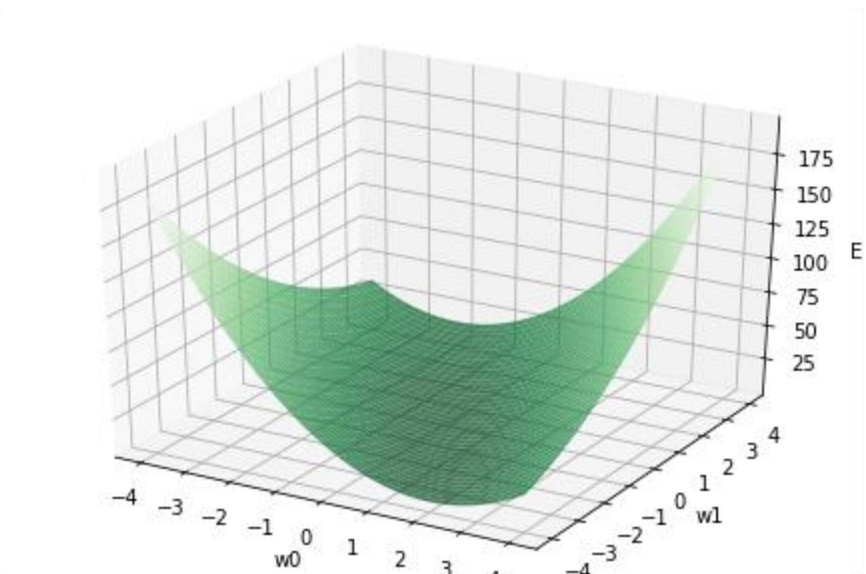
正則化項を導入した9次多項式回帰の係数

誤差関数の幾何学的解釈(1/2)

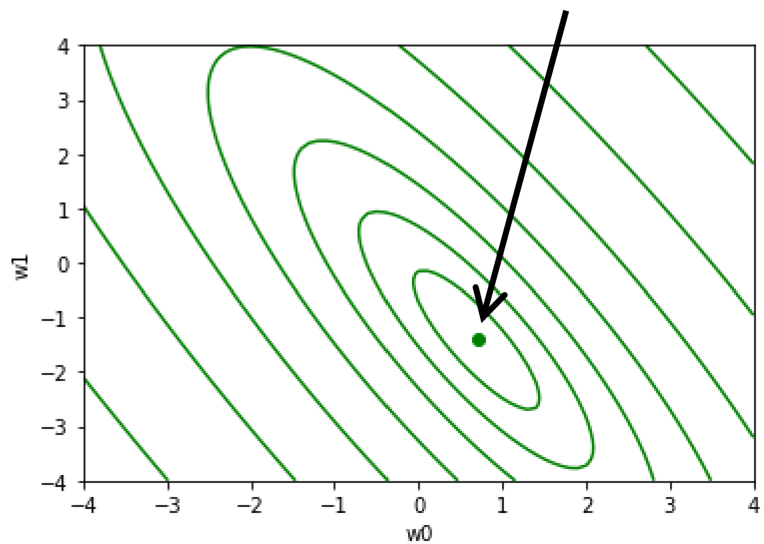
- ✓ 線形回帰の誤差関数を考える
- ✓ 係数w0, w1に対する誤差関数の等高線は楕円型となる

二乗和誤差項(線形回帰)

$$\sum_{i=1}^{n} \frac{1}{2}\left(y^{(i)} - \left(w_0 + w_1 x_1^{(i)}\right)\right)^2$$



二乗和誤差項が最小となる点



係数w0, w1に対する誤差項の変化

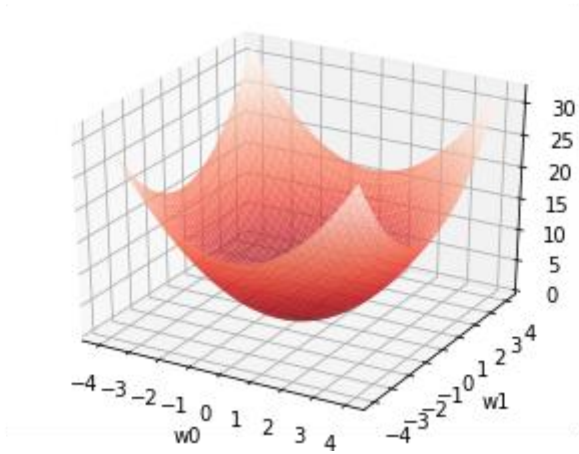
誤差関数の幾何学的解釈(2/2)

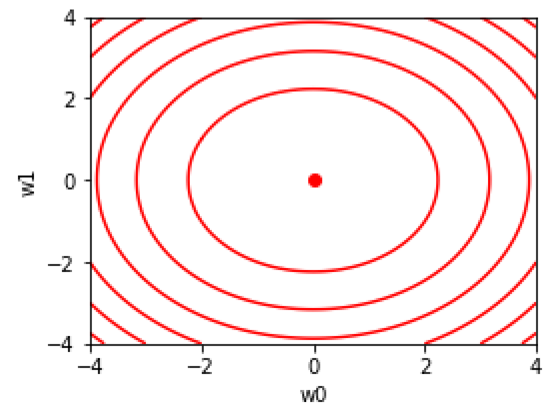
✓ 正則化項は下図のようになる

誤差関数が最小となる点は原点

正則化項(ridge)

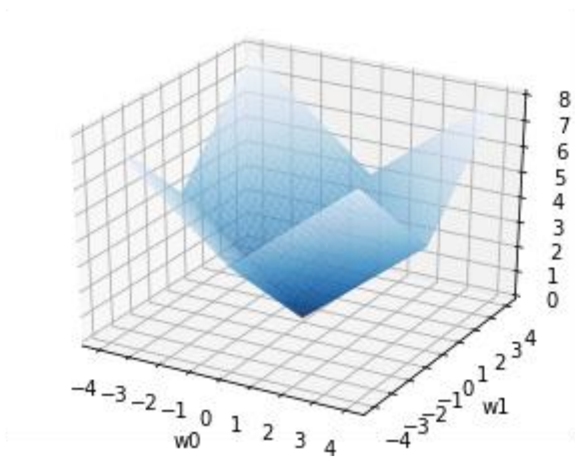
$$w_0{}^2 + w_1{}^2$$

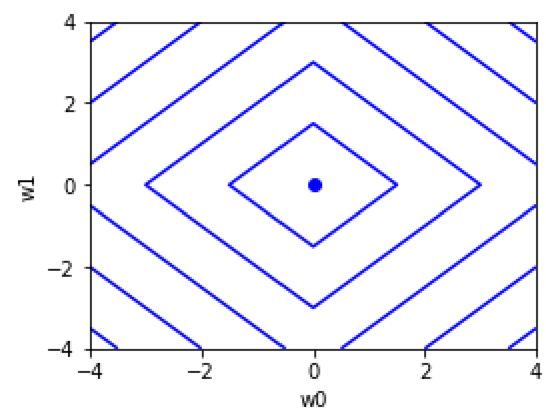




正則化項(lasso)

$$|w_0| + |w_1|$$





係数w0, w1に対する正則化項の変化

なぜlassoはスパース解が得られやすいのか(1/2)

✓ 正則化項付き誤差関数の最小化を制約あり最小化に変換

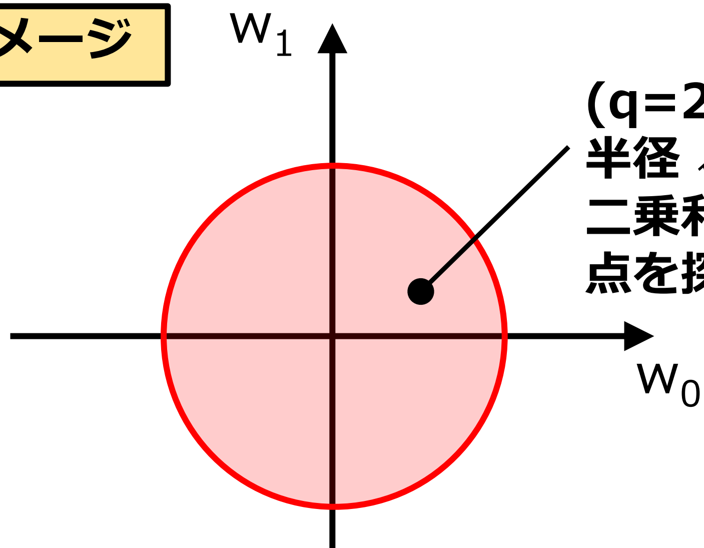
$$\min \quad \underbrace{\sum_{i=1}^{n} \frac{1}{2}\left(y^{(i)} - \left(w_0 + w_1 x_1^{(i)}\right)\right)^2}_{\text{二乗和誤差項}} + \frac{\alpha}{2} \underbrace{\sum_{i=0}^{2} |w_i|^q}_{\text{正則化項}}$$

ラグランジュの未定乗数法より等価

η: ラグランジュ乗数から決まる定数

$$\min \quad \underbrace{\sum_{i=1}^{n} \frac{1}{2}\left(y^{(i)} - \left(w_0 + w_1 x_1^{(i)}\right)\right)^2}_{\text{目的関数は二乗和誤差項のみ}} \quad s.t. \quad \underbrace{\sum_{i=0}^{2} |w_i|^q \le \eta}_{\text{正則化項は制約条件に}}$$

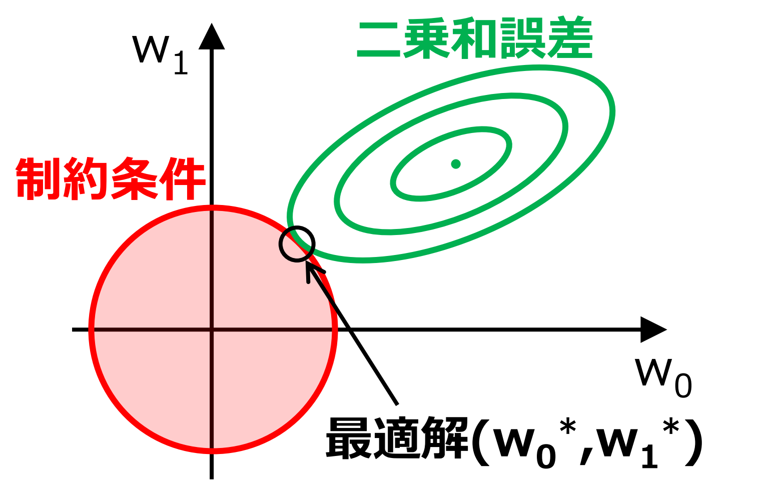
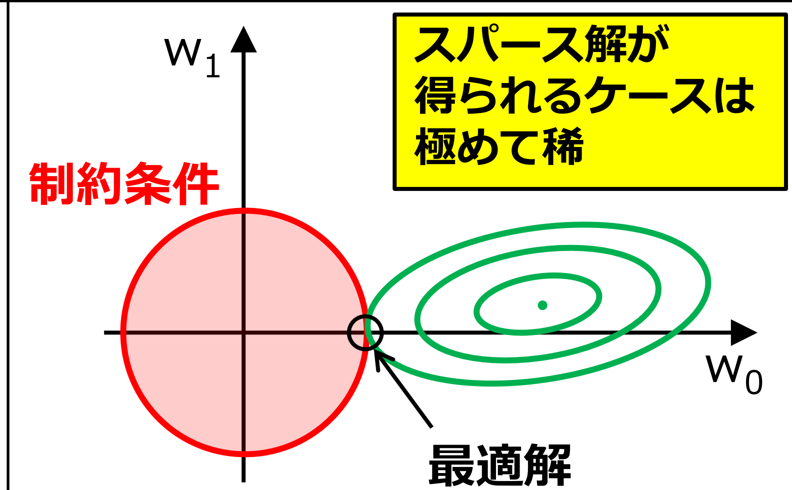
イメージ



(q=2の場合)
半径 √η の円内で
二乗和誤差が最小となる
点を探索する

なぜlassoはスパース解が得られやすいのか(2/2)

✓ lassoではスパース解となりやすい！

	スパース解が得られない場合	スパース解が得られる場合
ridge		
lasso