

Scikit-learn ゼミ

1.1.9. Bayesian Regression

1.1.9.1. Bayesian Ridge Regression

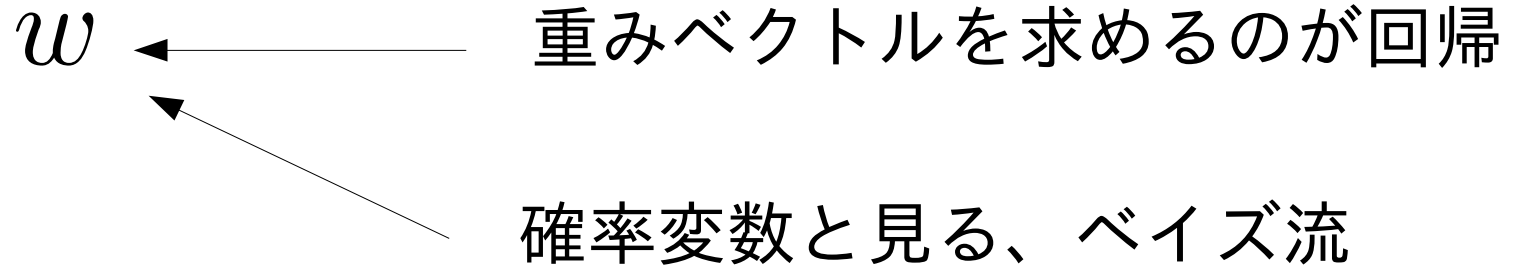
1.1.9.2. Automatic Relevance

Determination - ARD

新納浩幸

Bayesian Regression

回帰のベイズ版



長所

- ・ データにフィット
- ・ 正規化項の推定も同時にできる

短所

- ・ 計算が大変

式の結論

ビショップ本の 3.3 「ベイズ推定回帰」 参照

事前分布

$$p(\mathbf{w}) = N(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0)$$

事後分布

$$p(\mathbf{w} | \mathbf{t}) = N(\mathbf{w} | \mathbf{m}_N, \mathbf{S}_N)$$

$$\mathbf{m}_N = \mathbf{S}_N (\mathbf{S}_0^{-1} \mathbf{m}_0 + \beta \Phi^T \mathbf{t})$$

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta \Phi^T \Phi$$

計画行列

$$\Phi = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}$$

$\phi_j(\mathbf{x})$: 基底関数

ここでは線型を扱っているので、
 \mathbf{x} の次元が M 次元

$$\mathbf{x} = (x_0, x_1, \cdots, x_{M-1})$$

$$\phi_j(\mathbf{x}) = x_j$$

BayesianRidge

sklearn で提供されているベイズ線形回帰の関数

事前分布

$$p(\mathbf{w}) = N(\mathbf{w}|0, \lambda^{-1} \mathbf{I})$$

リッジ回帰

$$\min_{\mathbf{w}} (||\mathbf{X}\mathbf{w} - \mathbf{y}||^2 + \alpha ||\mathbf{w}||^2)$$



メタパラメータは λ と α

メタパラメータ α と λ

ベイズ流にこれらも一緒に推定する
事前分布にガンマ分布を利用

$$f(x) = \frac{x^{k-1} e^{-x/\theta}}{\Gamma(k)\theta^k}$$

← 更にパラメータ
 k と θ が存在

α に対する k と θ を α_1 と α_2 とおく
 λ に対する k と θ を λ_1 と λ_2 とおく

関数のデフォルト値として

$$\alpha_1 = \alpha_2 = \lambda_1 = \lambda_2 = 10^{-6}$$

実行例

```
>>> from sklearn import linear_model
>>> X = [[0., 0], [1., 1], [2., 2], [3., 3]]
>>> Y = [0., 1., 2., 3.]
>>> clf = linear_model.BayesianRidge()
>>> clf.fit(X, Y)
BayesianRidge(alpha_1=1e-06, alpha_2=1e-06, compute_score=False, copy_X=True,
              fit_intercept=True, lambda_1=1e-06, lambda_2=1e-06, n_iter=300,
              normalize=False, tol=0.001, verbose=False)
>>> clf.predict([[1., 0]])
array([ 0.50000013])
>>> clf.coef_
array([ 0.49999993,  0.49999993])
>>> clf.alpha_
1500001.0006260611
>>> clf.lambda_
1.999996264170399
>>>
```

推定した α と λ の値

Automatic Relevance Determination

ARD w の各次元の値の分散が異なるモデル

ARDRegression は BayesianRidge とほとんど同じ

$p(\mathbf{w}) = N(\mathbf{w}|0, \lambda^{-1} \mathbf{I})$ BayesianRidge の事前分布



$p(\mathbf{w}) = N(\mathbf{w}|0, A^{-1})$ ARDRegression の事前分布

$$\text{diag}(A) = \{\lambda_1, \lambda_2, \dots, \lambda_p\}$$

λ_i に対する k と θ はみな同じ λ_1 と λ_2