OKAYAMA UNIVERSITY

最尤推定

岡山大学 異分野基礎科学研究所

大槻純也

RIIS

前回の復習と今回の内容

前回の復習

線形回帰 (Linear regression)

$$y(x, \boldsymbol{w}) = \sum_{j=0}^{M} w_j \phi_j(x)$$

Basis function (nonlinear)

最小二乗法 (Least squares)

$$E(\boldsymbol{w}) = \frac{1}{2} \sum_{n=1}^{N} [y(x_n, \boldsymbol{w}) - t_n]^2$$

Error function

$$\boldsymbol{w}^* = (\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}\mathbf{t} \qquad \Phi_{nj} = \phi_j(x_n)$$

design matrix

Moore-Penrose pseudo inverse

正則化 (Regularization)

$$E(\boldsymbol{w}) = \frac{1}{2} \sum_{n=1}^{N} [y(x_n, \boldsymbol{w}) - t_n]^2 + \frac{\lambda}{2} \sum_{j} |w_j|^2$$

$$\boldsymbol{w}^* = (\Phi^{\mathrm{T}}\Phi + \lambda\mathbf{I})^{-1}\Phi^{\mathrm{T}}\mathbf{t}$$

regularization

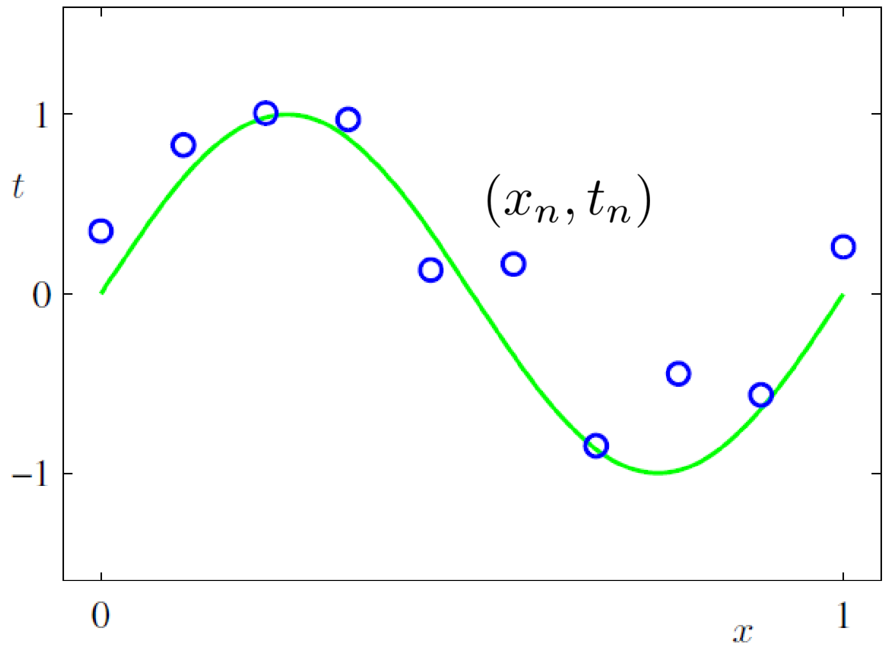
Training set

$$\mathbf{x} = \{x_1, x_2, \cdots, x_N\}$$

$$\mathbf{t} = \{t_1, t_2, \cdots, t_N\}$$

target

表記はPRMLに従う



PRML Fig. 1.2

今日の内容

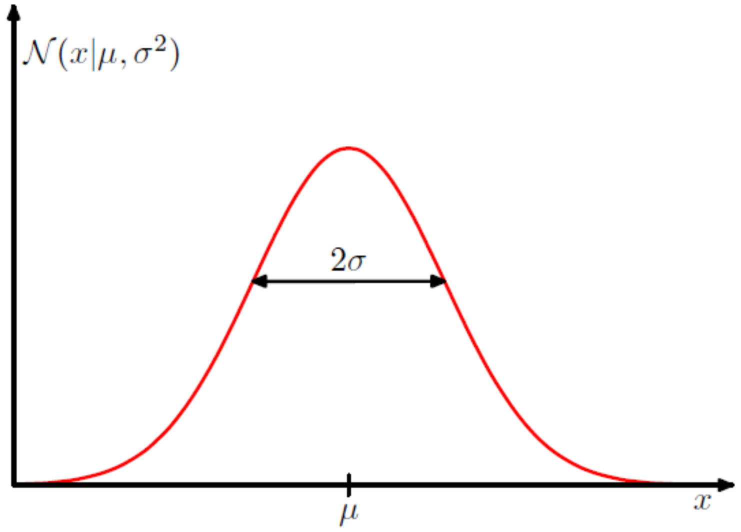
最小二乗法をベイズ推定の観点からみる

ガウス分布 (Gaussian distribution)

ガウス分布 または 正規分布 (Normal distribution)

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

誤差を表す際に用いられる。数学的に扱いやすい。



PRML Fig. 1.13

期待値 (Expectation value)　確率分布による平均値

$$\mathbb{E}[x] \equiv \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)x\,dx = \mu$$ mean (平均)

$$\mathbb{E}[x^2] \equiv \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)x^2\,dx = \mu^2 + \sigma^2$$

$$\mathrm{var}[x^2] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$ variance (分散)

σ はstandard deviation (標準偏差)

その他の性質

- Gaussian 2つの畳み込み積分はGaussian
- 多次元にも拡張可

尤度関数 (likelihood function)

線形回帰の前に、まずは、1変数 x を観測する問題を考える。

ガウス分布に従ってデータが生成されるとき、
値 x が観測される確率は

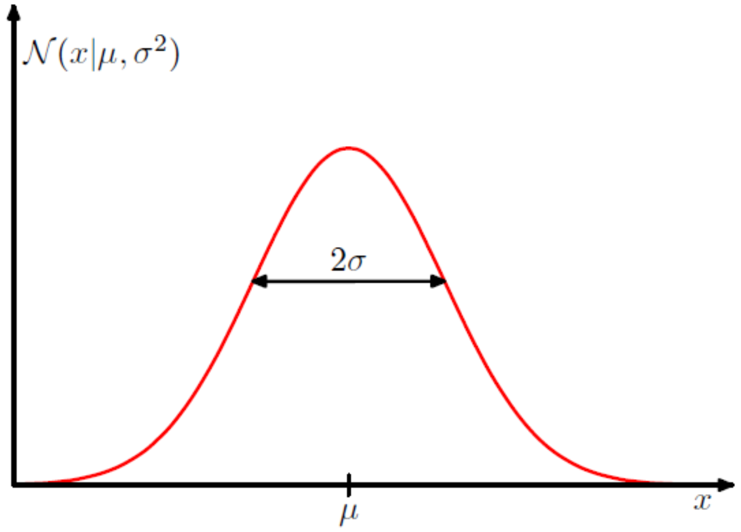
$$p(x|\mu,\sigma^2) = \mathcal{N}(x|\mu,\sigma^2)$$

x を観測したとする。そのデータの生成過程、
すなわち (μ,σ^2) は知らないとする。
$p(x|\mu,\sigma^2)$ を (μ,σ^2) の関数と見る。このとき、
$p(x|\mu,\sigma^2)$ を尤度関数 (likelihood function)と呼ぶ。

測定を N 回行った場合

$$\mathbf{x} = \{x_1, x_2, \cdots, x_N\}$$

$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2)$$



PRML Fig. 1.13

ベイズの定理

$$p(\mu,\sigma^2|\mathbf{x}) \propto p(\mathbf{x}|\mu,\sigma^2)p(\mu,\sigma^2)$$

知りたい量： (μ,σ^2)
観測データ： x

最尤推定 (Maximum likelihood)

尤度関数が最大になるように (μ, σ^2) を決定する

$$\ln p(\mathbf{x}|\mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

対数尤度 (log likelihood)

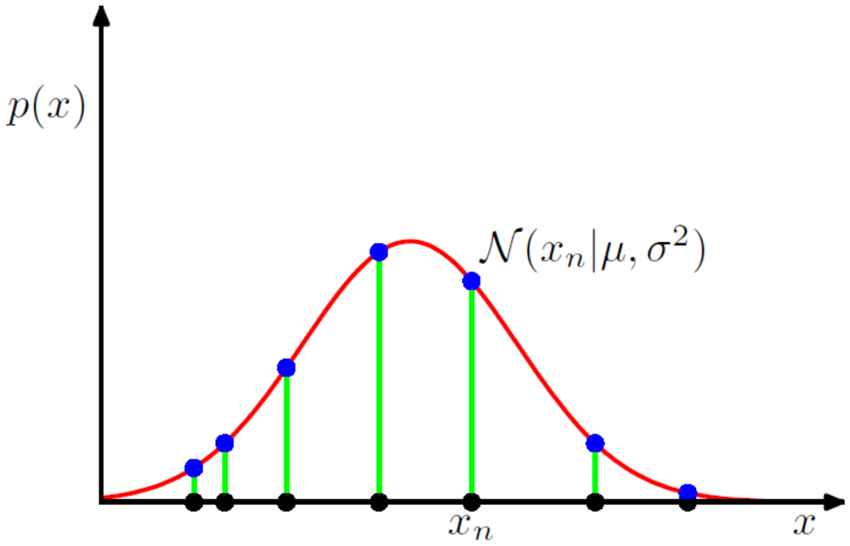
logは単調増加関数なので、対数尤度の微分でも解は同じ。logを取った方が計算が楽。

$\frac{\partial}{\partial\mu}\ln p(\mathbf{x}|\mu, \sigma^2) = 0$ より

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n$$ sample mean

$\frac{\partial}{\partial\sigma^2}\ln p(\mathbf{x}|\mu_{\mathrm{ML}}, \sigma^2) = 0$ より

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2$$ sample variance



PRML Fig. 1.14

ひとつだけ注意：MLは分散を過小評価する

$$\mathbb{E}[\mu_{\mathrm{ML}}] = \mu \qquad \boxed{\mathbb{E}[\sigma^2_{\mathrm{ML}}] = \frac{N-1}{N}\sigma^2}$$

導出はPRML Exercise 1.12

線形回帰

線形回帰に戻る

厳密な関数 $y(x, \boldsymbol{w})$ の誤差が乗ったデータが得られるとする

$$t = y(x, \boldsymbol{w}) + \epsilon$$ 誤差

$$p(\epsilon) = \mathcal{N}(\epsilon|0, \beta^{-1})$$ 誤差の確率分布

誤差が確率分布するので t も確率分布する

$$p(t|x, \boldsymbol{w}, \beta) = \mathcal{N}(t|y(x, \boldsymbol{w}), \beta^{-1})$$
$$= \sqrt{\frac{\beta}{2\pi}} \exp\left[-\frac{\beta}{2}(t - y(x, \boldsymbol{w}))^2\right]$$

$(\boldsymbol{w}, \beta)$ の関数とみる →尤度関数

$\beta = 1/\sigma^2$ は精度 (precision) パラメータと呼ばれる
σ^2 よりも β の方が都合が良いことが多い

N 個のデータ点が得られたとする

$$\mathbf{x} = \{x_1, x_2, \cdots, x_N\}$$
$$\mathbf{t} = \{t_1, t_2, \cdots, t_N\}$$

線形回帰の尤度関数

$$p(\mathbf{t}|\mathbf{x}, \boldsymbol{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, \boldsymbol{w}), \beta^{-1})$$
$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \exp\left[-\beta E(\boldsymbol{w})\right]$$

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N} [y(x_n, \boldsymbol{w}) - t_n]^2$$

統計力学のカノニカル分布と同じ形
統計力学の計算手法（モンテカルロ法など）が応用できそうなことが想像できる

線形回帰の最尤推定

対数尤度 (log likelihood) を最大化

$$\ln p(\mathbf{t}|\boldsymbol{w}, \beta) = -\beta E(\boldsymbol{w}) + \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi)$$

条件 $\mathbf{x}$ は省略

$$\frac{\partial}{\partial \boldsymbol{w}} \ln p(\mathbf{t}|\boldsymbol{w}, \beta) = 0$$ より

$$\boldsymbol{w}_{\mathrm{ML}} = \arg\min_{\boldsymbol{w}} E(\boldsymbol{w}) = (\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}\mathbf{t}$$

最小二乗法と一致

$$\frac{\partial}{\partial \beta} \ln p(\mathbf{t}|\boldsymbol{w}_{\mathrm{ML}}, \beta) = 0$$ より

$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{2}{N} E(\boldsymbol{w}_{\mathrm{ML}})$$

平均二乗誤差

t の予測分布 (Predictive distribution)

$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|y(x, \boldsymbol{w}_{\mathrm{ML}}), \beta_{\mathrm{ML}}^{-1})$$

新しい測定点

初めの仮定通りガウス分布
分散を過小評価する傾向がある点だけ注意

結局
最尤推定 ＝ 最小二乗法

MAP推定

ベイズの定理

事後確率分布　尤度関数　事前確率分布

$$p(\boldsymbol{w}|\mathbf{t},\beta) \propto p(\mathbf{t}|\boldsymbol{w},\beta)p(\boldsymbol{w})$$

知りたいのは $\boldsymbol{w}$

事後確率分布を最大化する

Maximum posterior (MAP) 推定
(最大事後確率推定)

事前確率として次の関数を考える

$$p(\boldsymbol{w}|\alpha) = \prod_j \mathcal{N}(w_j|0,\alpha^{-1})$$
$$= \left(\frac{\alpha}{2\pi}\right)^{\frac{M+1}{2}} \exp\left(-\frac{\alpha}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}\right)$$

すると、事後確率分布は

$$\ln p(\boldsymbol{w}|\mathbf{t},\alpha,\beta) = -\frac{\beta}{2}\sum_{n=1}^{N}[y(x_n,\boldsymbol{w})-t_n]^2 - \frac{\alpha}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + f(\alpha,\beta)$$

この最大化は正則化付き最小二乗法と等価

正則化パラメータ $\lambda = \alpha/\beta$

まとめ

ベイズ推定の言葉を使うと

最小二乗法 → 最尤推定
正則化付き最小二乗法 → MAP推定
正則化 → 事前分布