

# ベイズ線形回帰

岡山大学 異分野基礎科学研究所

大槻純也



# 前回の復習

座標  $x$  に依存した予測の不確定性を記述するためには  
パラメータ  $w$  の分布を考慮に入れる必要がある

= 最尤推定値  $w_{ML}$  まわりのゆらぎ

計算すべきは

ベイズ統計に基づく 予測分布 (Predictive distribution)

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int dw p(t|x, w) p(w|\mathbf{x}, \mathbf{t})$$

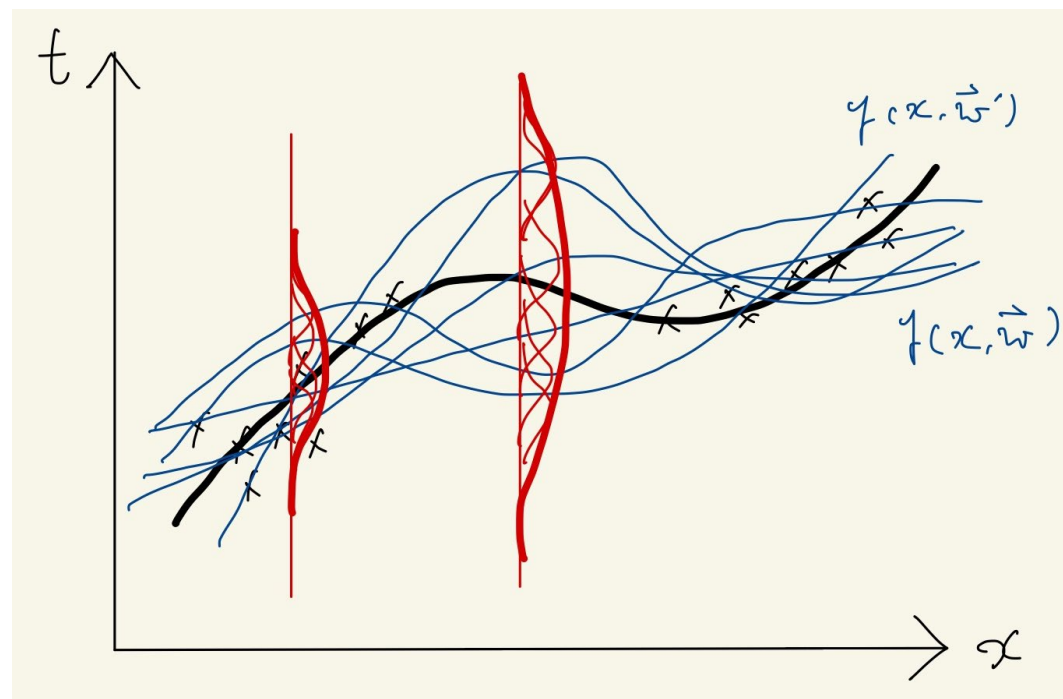
sum rule

パラメータ  $w$  の事後確率分布

パラメータ  $w$  が与えられたときの予測分布

パラメータ  $w$  の事後確率分布

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



# 必要な計算

ベイズ統計に基づく予測分布 (Predictive distribution)

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int d\mathbf{w} \underset{\textcircled{2}}{p(t|x, \mathbf{w})} \underset{\textcircled{3}}{p(\mathbf{w}|\mathbf{x}, \mathbf{t})} \quad \cdot \cdot \cdot \textcircled{1}$$

②  $\mathbf{w}$  が与えられたときの予測分布

$$p(t|x, \mathbf{w}) = \mathcal{N}(t|y(x, \mathbf{w}), \beta^{-1}) \propto \exp \left[ -\frac{\beta}{2} (t - \mathbf{w}^T \phi(x))^2 \right]$$

線形モデル

$$y(x, \mathbf{w}) = \mathbf{w}^T \phi(x)$$

③  $\mathbf{w}$  の事後確率分布

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}) = \prod_{n=1}^N \mathcal{N}(t_n|y(x_n, \mathbf{w}), \beta^{-1}) \prod_{j=1}^M \mathcal{N}(w_j|0, \alpha^{-1}) \\ \propto \exp \left[ -\frac{\beta}{2} (\mathbf{t} - \Phi \mathbf{w})^T (\mathbf{t} - \Phi \mathbf{w}) - \frac{\alpha}{2} \mathbf{w}^T \mathbf{w} \right]$$

Bayes' theorem

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

尤度関数    事前分布

結果はガウス分布 (ガウス分布の畳み込み積分もまたガウス分布)

$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x)) \quad \text{分散が } x \text{ に依存する}$$

今日やること

多変量ガウス分布 (multivariate Gaussian)  
の関係式を使って計算する

# 式③

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}) = \prod_{n=1}^N \mathcal{N}(t_n|y(x_n, \mathbf{w}), \beta^{-1}) \prod_{j=1}^M \mathcal{N}(w_j|0, \alpha^{-1})$$

1変数ガウス分布

線形モデル

$$y(x, \mathbf{w}) = \phi(x)^T \mathbf{w}$$

$$= \mathcal{N}(\mathbf{t}|\Phi\mathbf{w}, \beta^{-1}\mathbf{I})\mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

多変量ガウス分布

$$\propto \exp \left[ -\frac{\beta}{2}(\mathbf{t} - \Phi\mathbf{w})^T(\mathbf{t} - \Phi\mathbf{w}) - \frac{\alpha}{2}\mathbf{w}^T\mathbf{w} \right]$$

計画行列

$$\Phi_{nj} = \phi_j(x_n)$$

$N \times M$  行列

$\mathbf{w}$  に関して平方完成すると

これは公式を使わなくても計算できる

しかし公式を使いこなすのもテクニックのひとつ

対応する公式は、PRML 式(2.116)

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{w}_{\text{ML}}, S)$$

$$\mathbf{w}_{\text{ML}} = \beta S \Phi^T \mathbf{t}$$

最尤推定におけるパラメータの最適値

$$S = (\beta \Phi^T \Phi + \alpha \mathbf{I})^{-1}$$

最尤推定における共分散行列  
(最適値周りの揺らぎ)

# 式①

$$\begin{aligned} p(t|x, \mathbf{x}, \mathbf{t}) &= \int d\mathbf{w} p(t|x, \mathbf{w}) p(\mathbf{w}|\mathbf{x}, \mathbf{t}) \\ &= \int d\mathbf{w} \mathcal{N}(t|\phi(x)^T \mathbf{w}, \beta^{-1}) \mathcal{N}(\mathbf{w}|\mathbf{w}_{\text{ML}}, S) \\ &\propto \int d\mathbf{w} \exp \left[ -\frac{\beta}{2} (t - \phi(x)^T \mathbf{w})^2 - \frac{1}{2} (\mathbf{w} - \mathbf{w}_{\text{ML}})^T S^{-1} (\mathbf{w} - \mathbf{w}_{\text{ML}}) \right] \end{aligned}$$

$\mathbf{w}$  に関して平方完成して積分すると

$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x))$$

$$m(x) = \beta \phi(x)^T S \Phi^T \mathbf{t}$$

$$s^2(x) = \beta^{-1} + \phi(x)^T S \phi(x)$$

この計算は結構大変

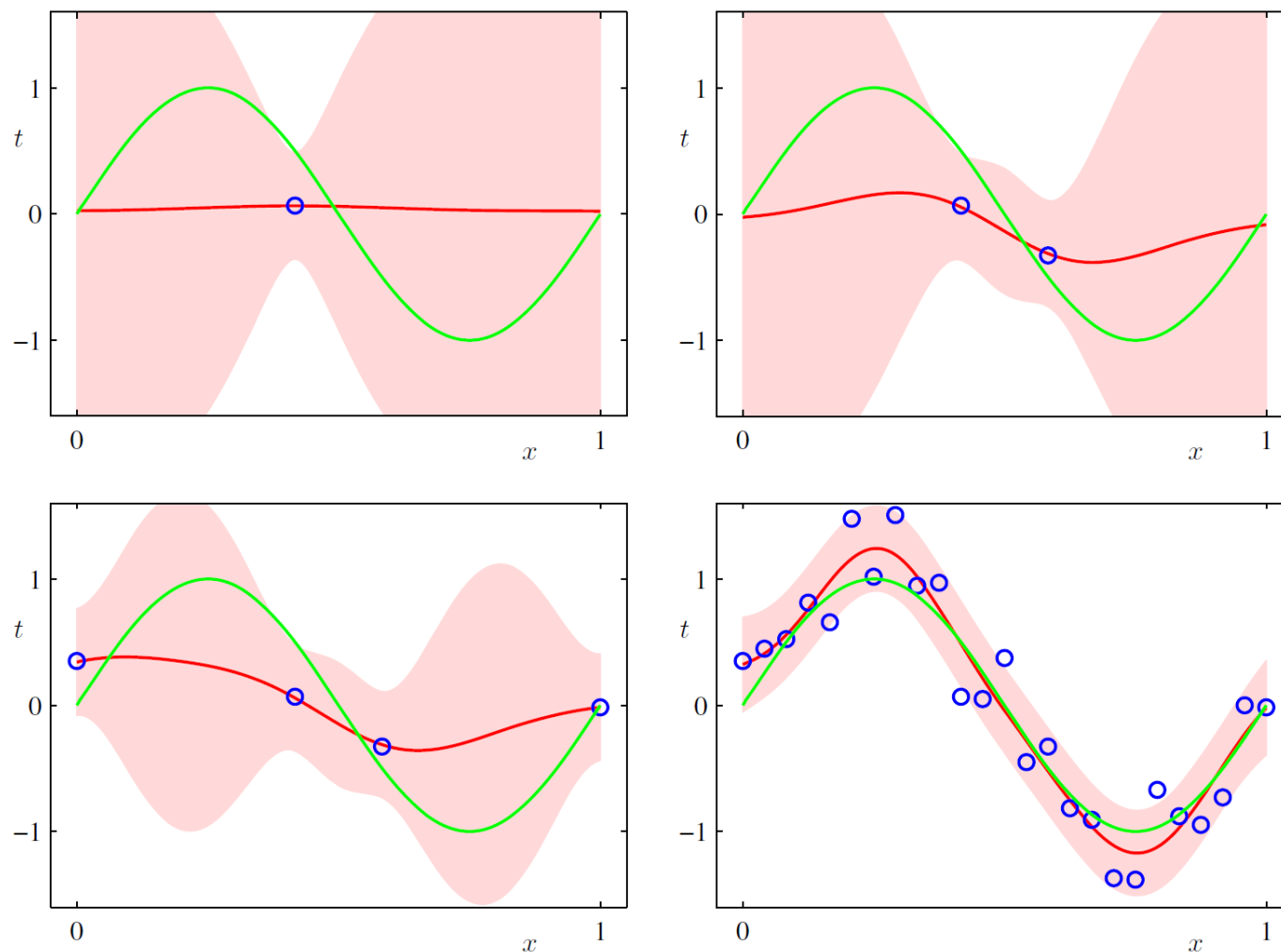
腕試しに計算してみるのもよい

公式を使う場合は、PRML 式(2.115)

$\mathbf{w}$  の分布による予測分布の拡がり

これが最尤推定と正しいベイズ推定の違い

# 予測分布



$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x))$$

緑：答え

青：測定データ

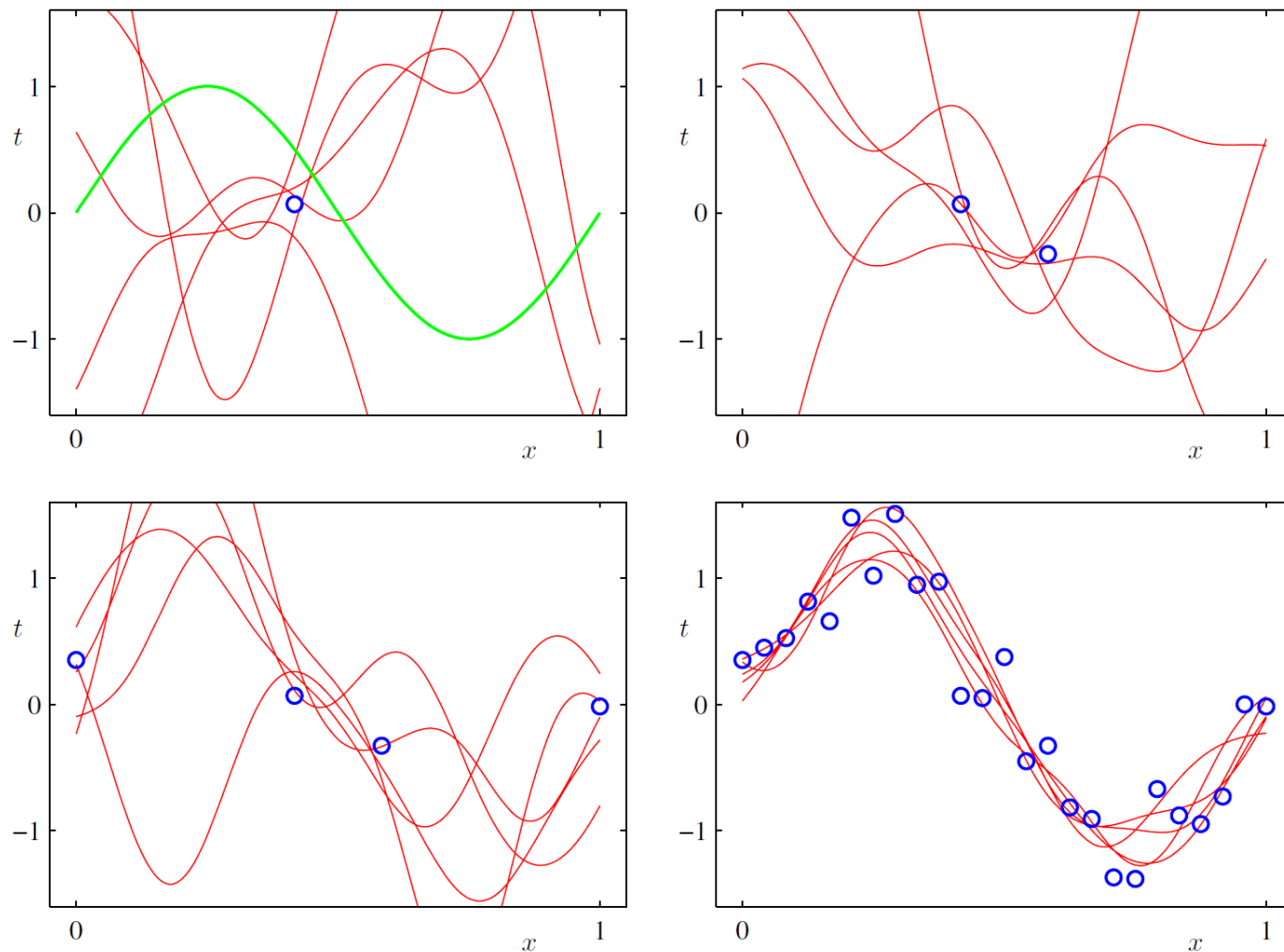
赤： $m(x)$

ピンク： $[m(x) - s(x), m(x) + s(x)]$

基底関数：  
ガウス関数（あとで）

PRML **Figure 3.8** Examples of the predictive distribution (3.58) for a model consisting of 9 Gaussian basis functions of the form (3.4) using the synthetic sinusoidal data set of Section 1.1. See the text for a detailed discussion.

# 予測曲線



赤線：  
予測曲線の例

$$y(x, \mathbf{w}) = \phi^T \mathbf{w}$$

パラメータ  $\mathbf{w}$  の事後確率分布に  
従ってサンプルを抽出

$$p(\mathbf{w} | \mathbf{x}, \mathbf{t})$$

実際の予測分布は、赤線を中心とした  
分散  $\beta^{-1}$  のガウス分布  
それを積分したものが前頁の図

PRML **Figure 3.9** Plots of the function  $y(x, \mathbf{w})$  using samples from the posterior distributions over  $\mathbf{w}$  corresponding to the plots in Figure 3.8.

# 基底関数

1次元データ( $x, t$ )

$$y(x, \mathbf{w}) = \sum_j w_j \phi_j(x)$$

基底関数の例

多項式  $\phi_j(x) = x^j$

スプライン (区分多項式)

ガウス関数

$$\phi_j(x) = \exp \left[ -\frac{(x - \mu_j)^2}{2s^2} \right]$$

シグモイド (sigmoid) 関数

$$\phi_j(x) = \sigma \left( \frac{x - \mu_j}{s} \right)$$

$$\sigma(a) = \frac{1}{1 + e^{-a}} \quad \text{logistic関数}$$

## 基底関数

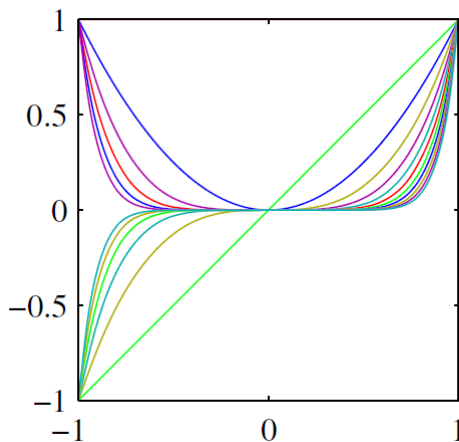
空間的に局在している関数が好ましい  
関数が広がりを持っていると、一つのデータ点が全ての  $w_j$  に影響する  
その意味で、多項式はあまりよくない

ちなみに、これまでは入力が1次元としてきたが、実際は何次元でも定式化は変わらない。

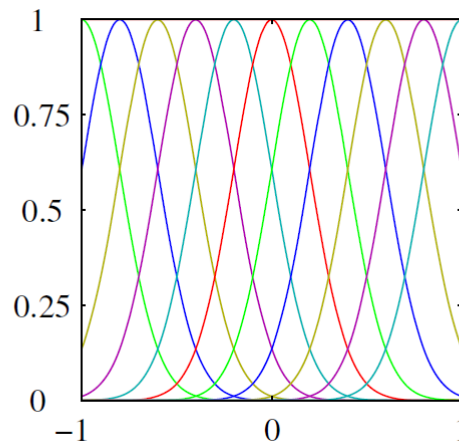
$$\mathbf{x} = (x_1, x_2, \dots, x_D)$$

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^M w_j \phi_j(\mathbf{x}) \quad \phi_j(x) \rightarrow \phi_j(\mathbf{x}) \text{ と置き換えるだけ}$$

多項式



ガウス関数



シグモイド関数

