東京都立大学集中講義「超伝導理工学特別講義」
第3回　2022年8月2～4日

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

ベイズの定理

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

事後確率分布　　尤度関数　　事前確率分布

Training set

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

Maximum posterior (MAP) 推定
$p(\boldsymbol{w}) = \text{const}$ なら最尤推定

$$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} p(\boldsymbol{w}|\mathbf{x},\mathbf{t})$$

➔(正則化付き)
最小二乗法に一致

MAP推定は…

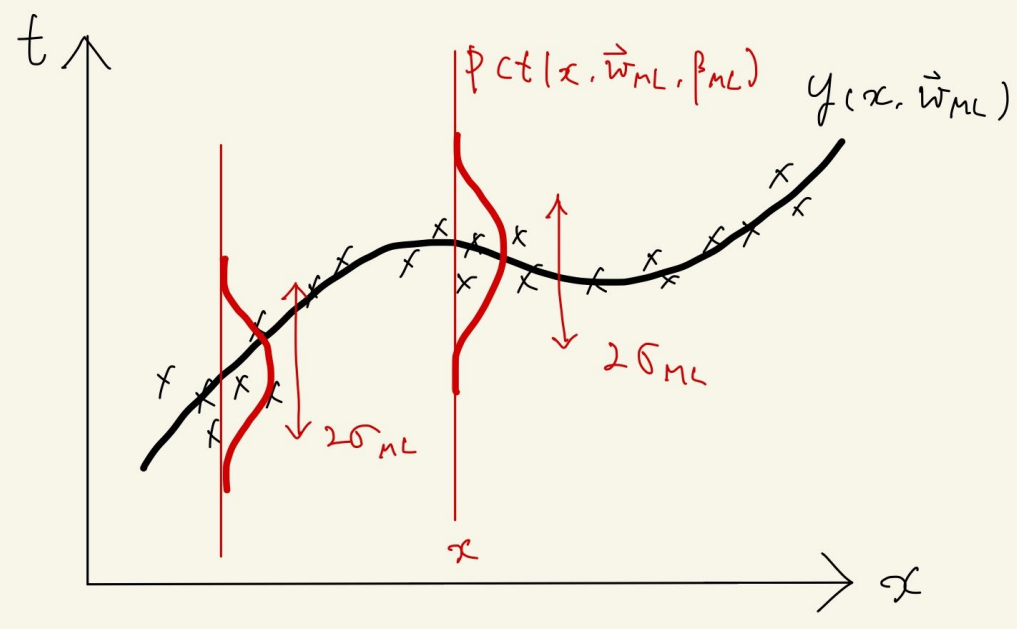
パラメータ $\boldsymbol{w}$ の分布を無視する近似

物理の言葉を使うと
- パラメータ $\boldsymbol{w}$ のゆらぎを無視する近似
- 平均場近似、鞍点近似

予測分布 (Predictive distribution)

最尤推定の予測分布 (Predictive distribution)

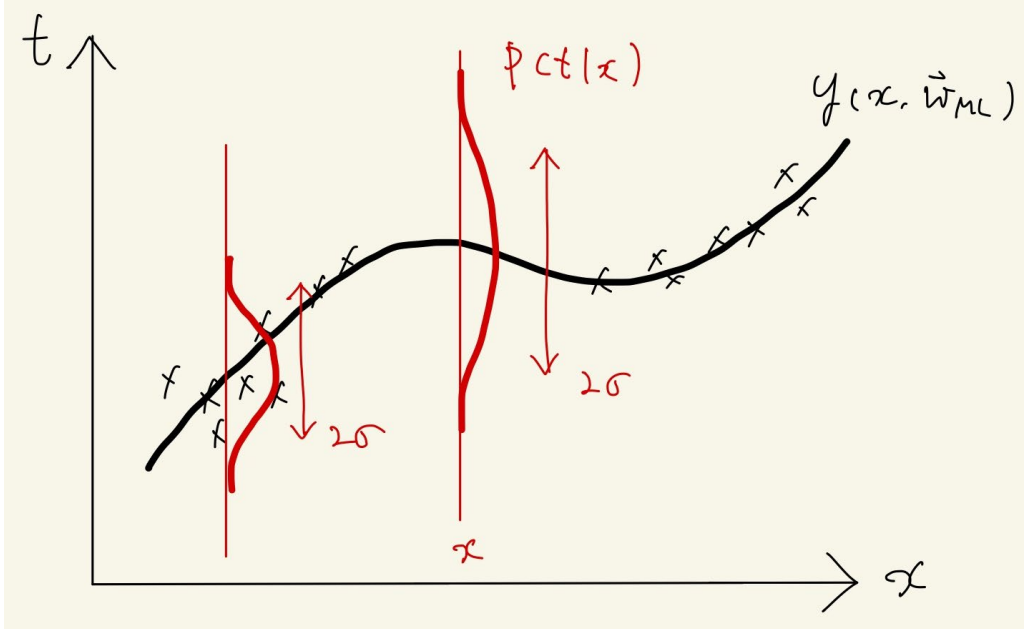
$$p(t|x, \boldsymbol{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}) = \mathcal{N}(t|y(x, \boldsymbol{w}_{\mathrm{ML}}), \beta_{\mathrm{ML}}^{-1})$$



σ は x に依らない

$$\sigma = \sigma_{\mathrm{ML}} = \beta_{\mathrm{ML}}^{-1/2}$$

しかしこのように、
データ点が一様に存在しない場合



σ は x に依存すべき

$$\sigma = \sigma(x)$$

やりたいこと：ベイズ線形回帰

パラメータ $\boldsymbol{w}$ の分布を考慮に入れる

＝ 最尤推定値 $\boldsymbol{w}_{\mathrm{ML}}$ まわりのゆらぎ

$\boldsymbol{w}$ の事後確率分布

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

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

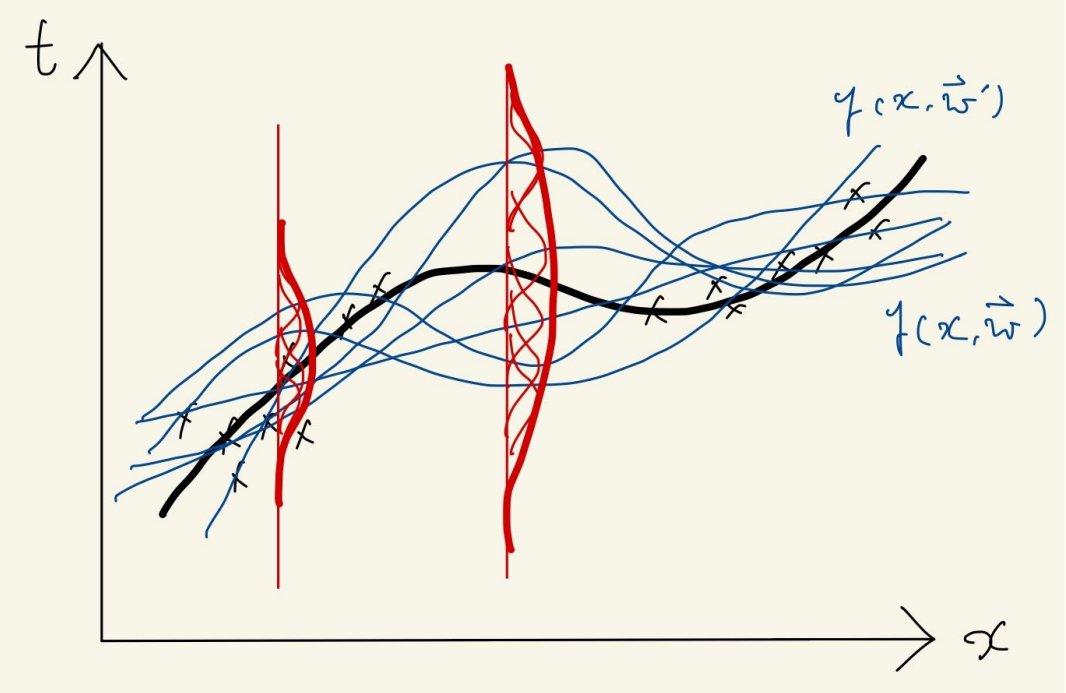
$$p(t|x,\mathbf{x},\mathbf{t}) = \int d\boldsymbol{w} p(t|x,\boldsymbol{w})p(\boldsymbol{w}|\mathbf{x},\mathbf{t})$$

sum rule

パラメータ $\boldsymbol{w}$ の事後確率分布

パラメータ $\boldsymbol{w}$ が与えられたときの予測分布

座標 x に依存した予測の不確定性を記述する



必要な計算

予測分布 (Predictive distribution)

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

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

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

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

Bayes' theorem

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

Gaussianの積の積分もまたGaussianになる

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

分散が x に依存する

これを導出するには
多変量ガウス分布 (multivariate Gaussian)
の関係式が必要

多変量ガウス分布 (multivariate Gaussian distribution)

OKAYAMA UNIVERSITY

1次元のガウス分布

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

μ : mean (平均)
σ^2 : variance (分散)

D 次元のガウス分布

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

$\boldsymbol{\mu}$: mean (平均)
$\boldsymbol{\Sigma}$: covariance matrix (共分散行列)
$|\boldsymbol{\Sigma}|$: determinant (行列式)

$$\mathbb{E}[\mathbf{x}]=\int\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\mathbf{x}d\mathbf{x}=\boldsymbol{\mu}$$ mean

$$\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathrm{T}}]=\int\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\mathbf{x}\mathbf{x}^{\mathrm{T}}d\mathbf{x}=\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}+\boldsymbol{\Sigma}$$

$$\mathrm{cov}[\mathbf{x}]=\mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}]=\boldsymbol{\Sigma}$$ covariance

Covariance（共分散)

Covariance matrix $\mathbf{\Sigma}$ の性質

◆ Real symmetric (対称行列)

$$\Sigma_{ij} = \Sigma_{ji}$$ よって固有値は実数

◆ Positive definite (正定値)
全ての固有値が正

$$\mathbf{\Sigma}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \quad \lambda_i > 0$$

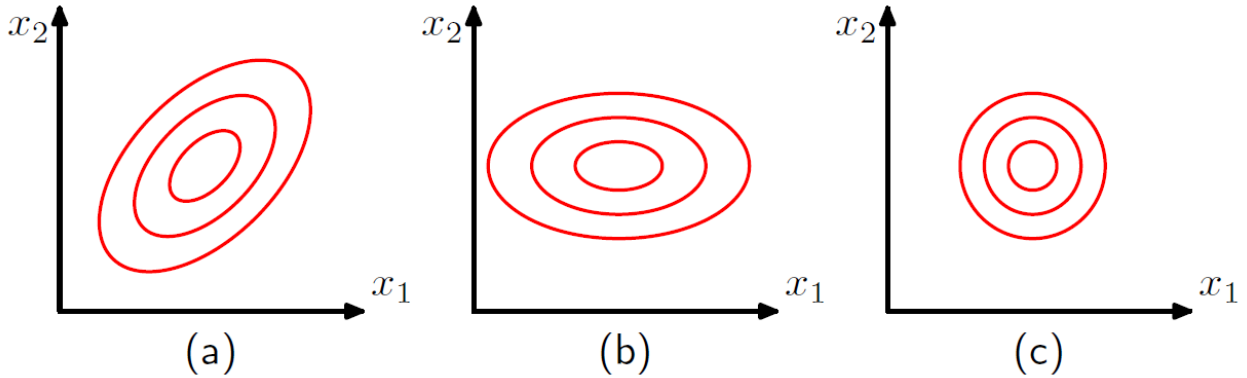
なぜなら

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}) = \prod_{i=1}^{D} \frac{1}{(2\pi\lambda_i)^{D/2}} \exp\left(-\frac{y_i^2}{2\lambda_i}\right)$$

$$y_i = \boldsymbol{u}_i^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu})$$

$$\mathbf{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \qquad \mathbf{\Sigma} = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \qquad \mathbf{\Sigma} \propto \mathbf{I}$$

$\Sigma_{12} > 0$ $\Sigma_{11} > \Sigma_{22}$



PRML Fig. 2.8

正の相関

x_1 と x_2 が近い値を取りやすい

相関なし

x_1 と x_2 の分布の幅が異なる

相関なし

x_1 と x_2 が同じ分布し従う

準備

D 次元ガウス分布を考え、
$\mathbf{x}$ を M 成分と $D-M$ 成分に分割する

$$p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$$

$\boldsymbol{\Sigma}$ は対称行列なので $\boldsymbol{\Sigma}_{ab} = \boldsymbol{\Sigma}_{ba}^{\mathrm{T}}$

Precision matrix $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ を定義

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix}$$

$\boldsymbol{\Sigma}$ と $\boldsymbol{\Lambda}$ の関係

$$\boldsymbol{\Lambda}_{aa} = (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}$$
$$\boldsymbol{\Lambda}_{ab} = -(\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}$$

計算する量によって $\boldsymbol{\Sigma}$ と $\boldsymbol{\Lambda}$ のどちらで表記がシンプルになる場合がある。

Marginal Gaussian distribution

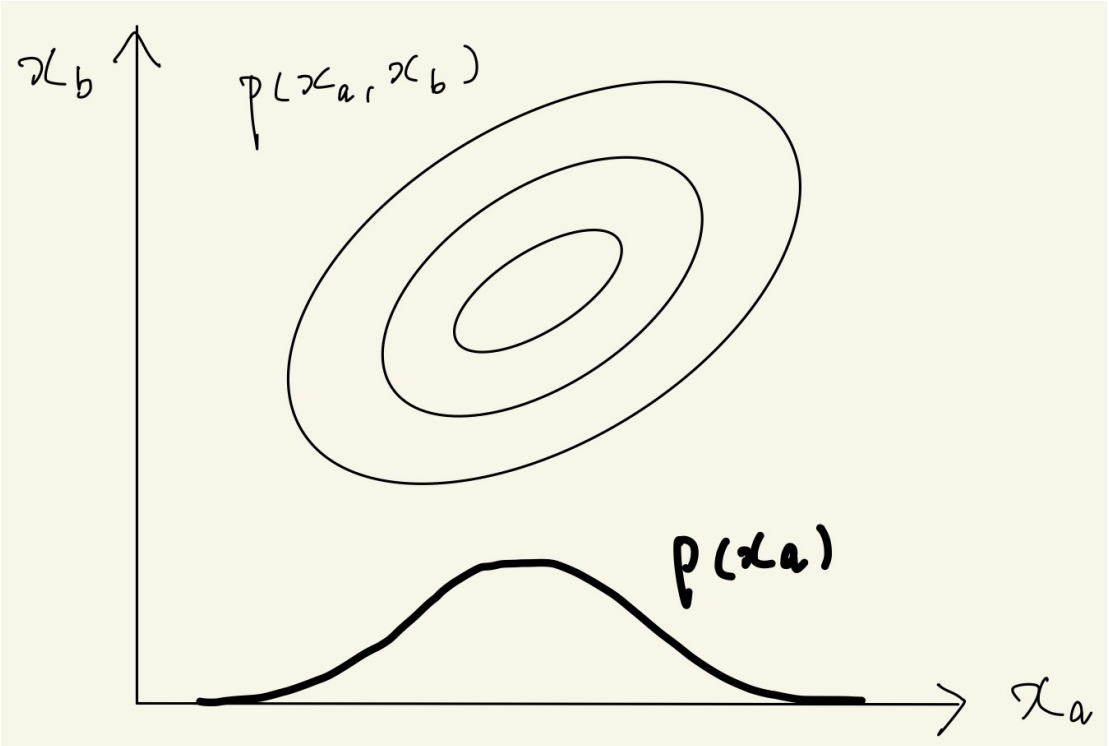
変数 $\mathbf{x}_b$ について周辺化（積分）する

$$p(\mathbf{x}_a) = \int p(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_b$$

結果は、$\boldsymbol{\mu}$ や $\boldsymbol{\Sigma}$ からaa成分のみを
取り出したものになる

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a | \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$$

$\boldsymbol{\Lambda}$ ではなく $\boldsymbol{\Sigma}$ でaa成分を抜き出す点に注意。
結果はシンプルだが、証明は意外と大変
（PRML §2.3.2）



周辺分布は積分に対応
Gaussianを積分してもGaussian

Conditional Gaussian distribution

条件付き確率を得るには

$$p(\mathbf{x}_a, \mathbf{x}_b) = p(\mathbf{x}_a|\mathbf{x}_b)p(\mathbf{x}_b)$$

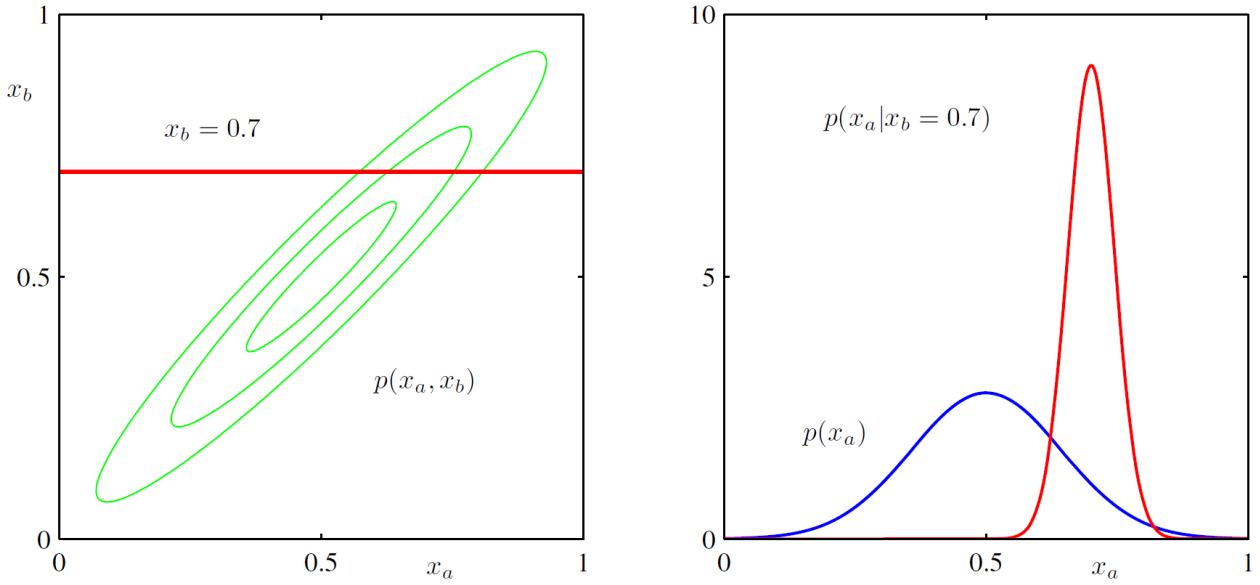
product rule

の両辺を比較する。結果は

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b})$$

$$\begin{aligned}\boldsymbol{\mu}_{a|b} &= \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\ &= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b)\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{a|b} &= \boldsymbol{\Lambda}_{aa}^{-1} \\ &= \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}\end{aligned}$$



PRML Fig. 2.9

条件付き確率はカットに対応
Gaussianをカットしてもガウシアン

公式集

実際に必要になったら公式集を見るのがよい (Bishop, PRML)
どこに載っているかを覚えておくことが重要

2.3. The Gaussian Distribution

2.3.1 Conditional Gaussian distributions

2.3.2 Marginal Gaussian distributions

Partitioned Gaussians

Given a joint Gaussian distribution $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Lambda} \equiv \boldsymbol{\Sigma}^{-1}$ and

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \tag{2.94}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix}. \tag{2.95}$$

Conditional distribution:

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1}) \tag{2.96}$$

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b). \tag{2.97}$$

Marginal distribution:

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa}). \tag{2.98}$$

2.3.3 Bayes' theorem for Gaussian variables

Marginal and Conditional Gaussians

Given a marginal Gaussian distribution for $\mathbf{x}$ and a conditional Gaussian distribution for $\mathbf{y}$ given $\mathbf{x}$ in the form

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \tag{2.113}$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}) \tag{2.114}$$

the marginal distribution of $\mathbf{y}$ and the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ are given by

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}) \tag{2.115}$$

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma}) \tag{2.116}$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A})^{-1}. \tag{2.117}$$

ここまでは準備。いよいよ計算

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

$$p(t|x,\mathbf{x},\mathbf{t}) = \int d\boldsymbol{w} p(t|x,\boldsymbol{w}) p(\boldsymbol{w}|\mathbf{x},\mathbf{t}) \quad \cdots ①$$

② ③

② $\boldsymbol{w}$ が与えられたときの予測分布

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

③ $\boldsymbol{w}$ の事後確率分布

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

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

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)

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

分散が x に依存する

線形モデル

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

Bayes' theorem

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

尤度関数　事前分布

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

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

1変数ガウス分布

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

多変量ガウス分布

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

線形モデル

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

計画行列

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

$N \times M$ 行列

$\boldsymbol{w}$ に関して平方完成すると

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

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)

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

最尤推定におけるパラメータの最適値

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

最尤推定における共分散行列

(最適値周りの揺らぎ)

式①

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

$\boldsymbol{w}$ に関して平方完成して積分すると

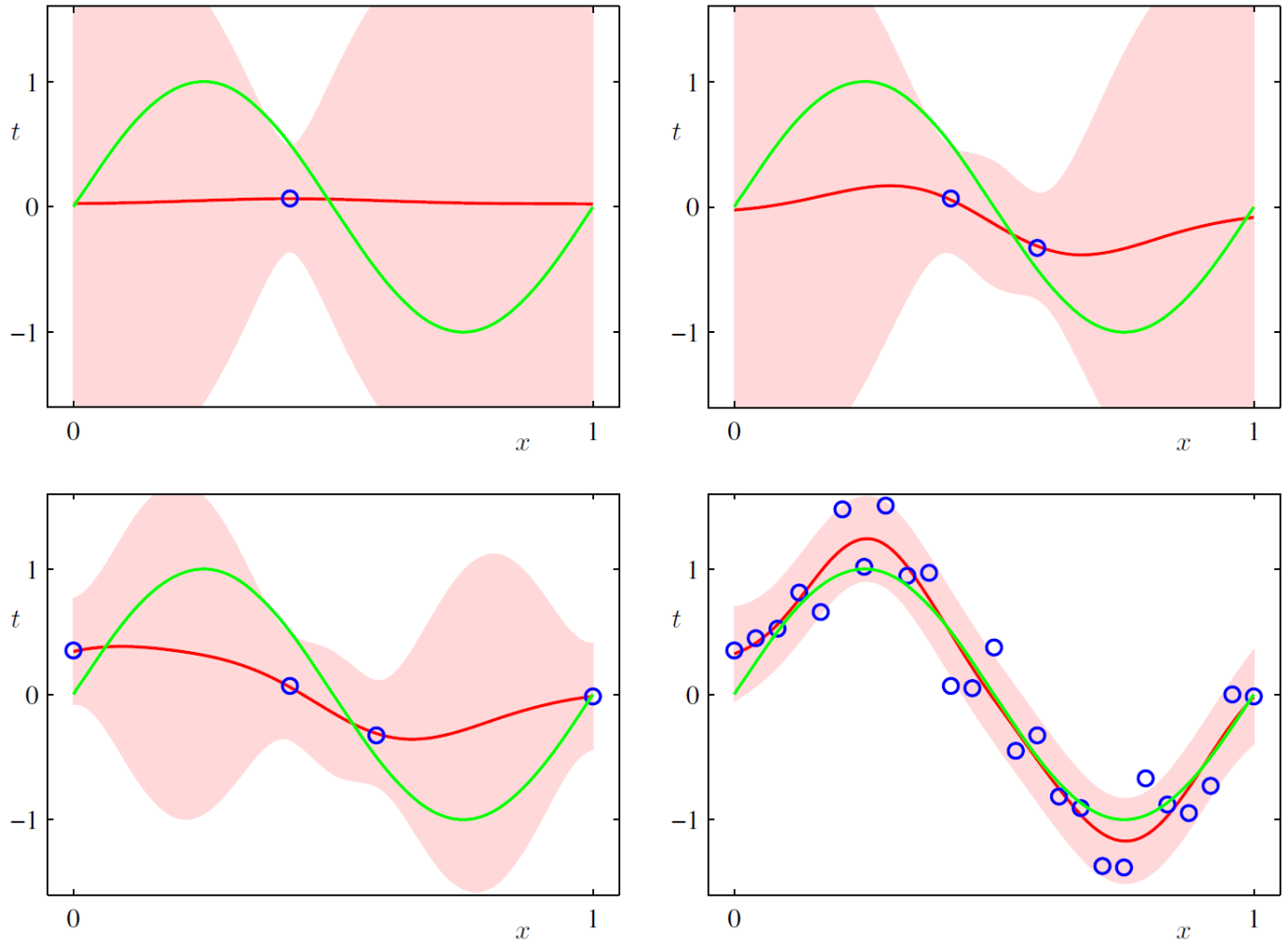
$$
\begin{aligned}
p(t|x,\mathbf{x},\mathbf{t}) &= \mathcal{N}(t|m(x), s^2(x)) \\
m(x) &= \beta \boldsymbol{\phi}(x)^{\mathrm{T}} S \Phi^{\mathrm{T}} \mathbf{t} \\
s^2(x) &= \beta^{-1} + \boldsymbol{\phi}(x)^{\mathrm{T}} S \boldsymbol{\phi}(x)
\end{aligned}
$$

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)

$\boldsymbol{w}$ 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



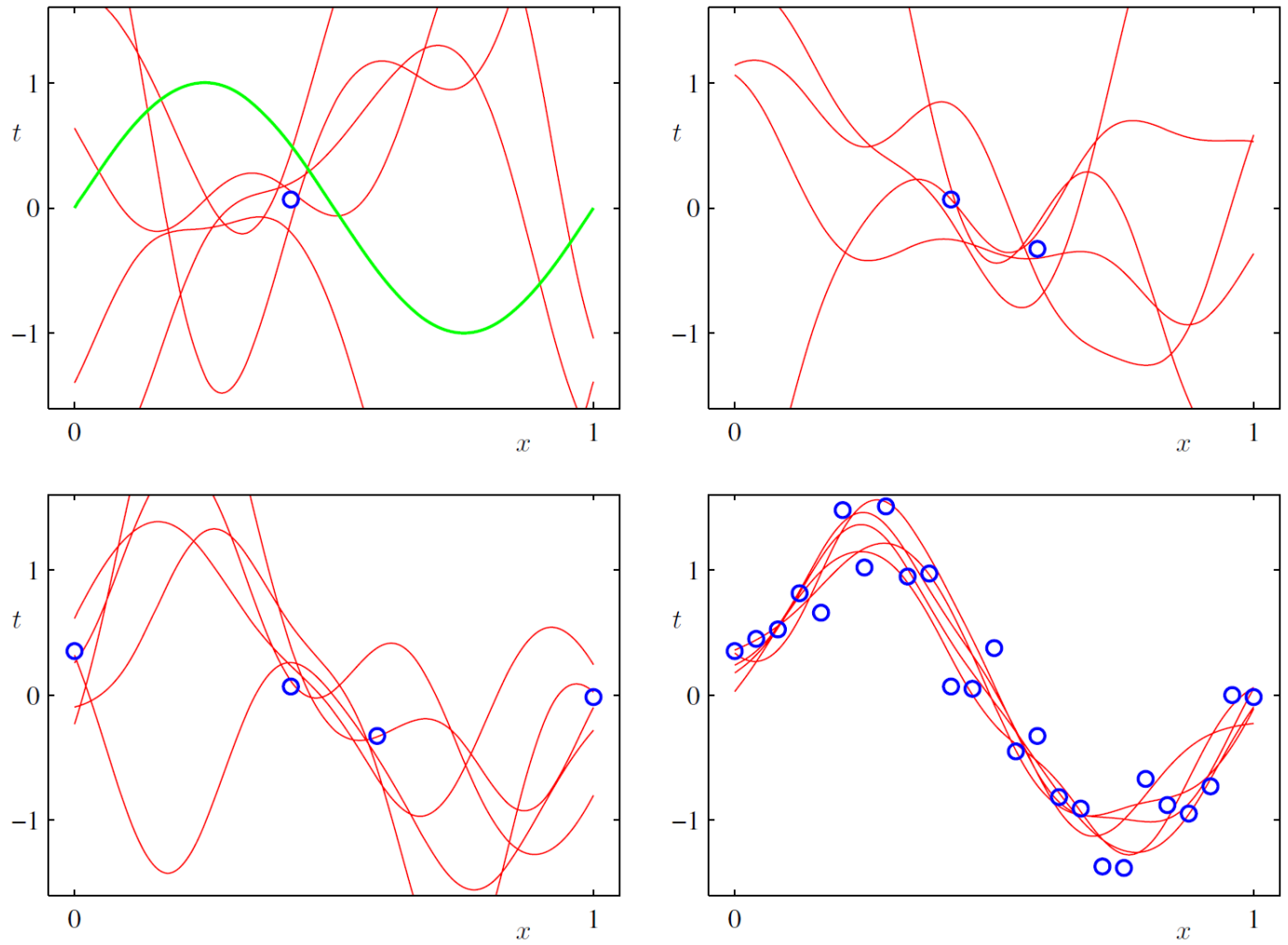
$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x))$$

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

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

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

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

実際の予測分布は、赤線を中心とした
分散 β^{-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.

まとめ

ベイズ線形回帰による予測分布

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

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

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

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

最尤推定における予測曲線

$$m(x) = \boldsymbol{\phi}(x)^{\mathrm{T}} \boldsymbol{w}_{\mathrm{ML}} \qquad \boldsymbol{w}_{\mathrm{ML}} = \beta S \Phi^{\mathrm{T}} \mathbf{t}$$

$\boldsymbol{w}$ 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

最尤推定における共分散行列

(最適値周りの揺らぎ)

補足：基底関数

1次元データ(x, t)

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

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

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

$$\sigma(a) = \frac{1}{1+e^{-a}}$$

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

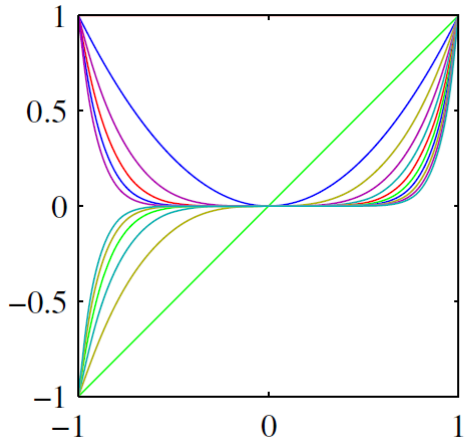
ちなみに、これまでは入力が1次元としてきたが、実際は何次元でも定式化は変わらない。

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

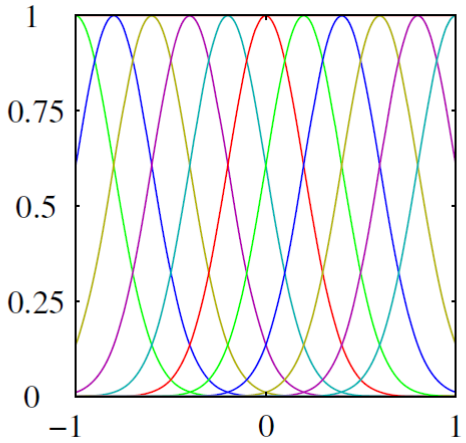
$$y(\mathbf{x}, \boldsymbol{w}) = \sum_{j=1}^{M} w_j \phi_j(\mathbf{x})$$

$\phi_j(x) \to \phi_j(\mathbf{x})$ と置き換えるだけ

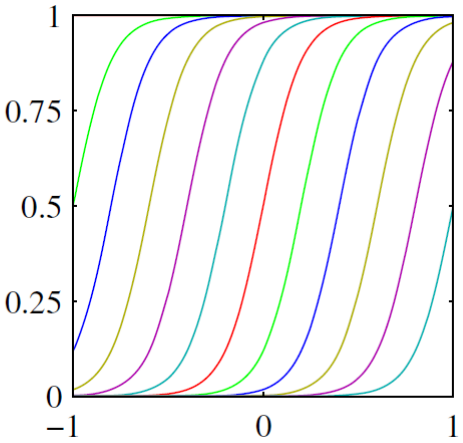
多項式



ガウス関数



シグモイド関数



PRML Fig. 3.1