

ベイズ線形回帰の準備： 多変量ガウス分布

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

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

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

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

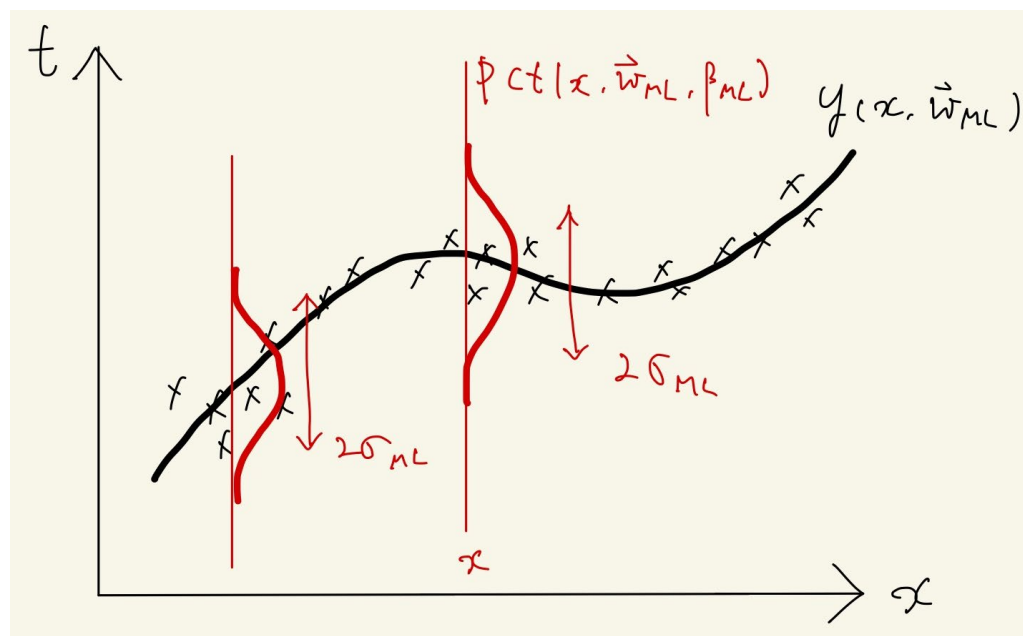
物理の言葉を使うと

- 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)

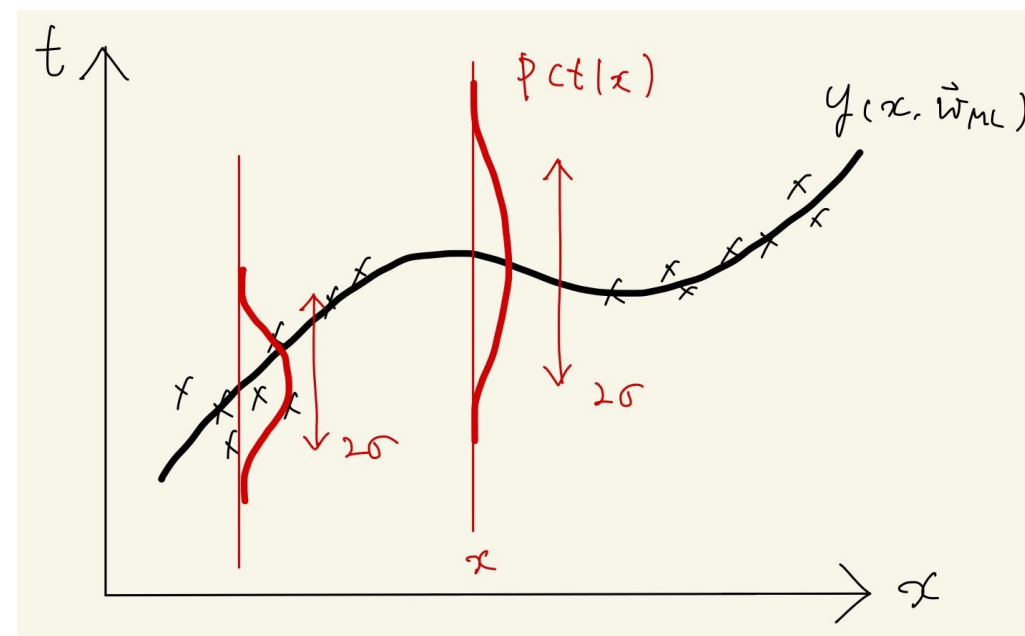
$$p(t|x, \mathbf{w}_{\text{ML}}, \beta_{\text{ML}}) = \mathcal{N}(t|y(x, \mathbf{w}_{\text{ML}}), \beta_{\text{ML}}^{-1})$$



σ は x に依らない

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

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 w の分布を考慮に入れる

パラメータ w の事後確率分布

$$p(w|x, t) \propto p(t|x, w)p(w)$$

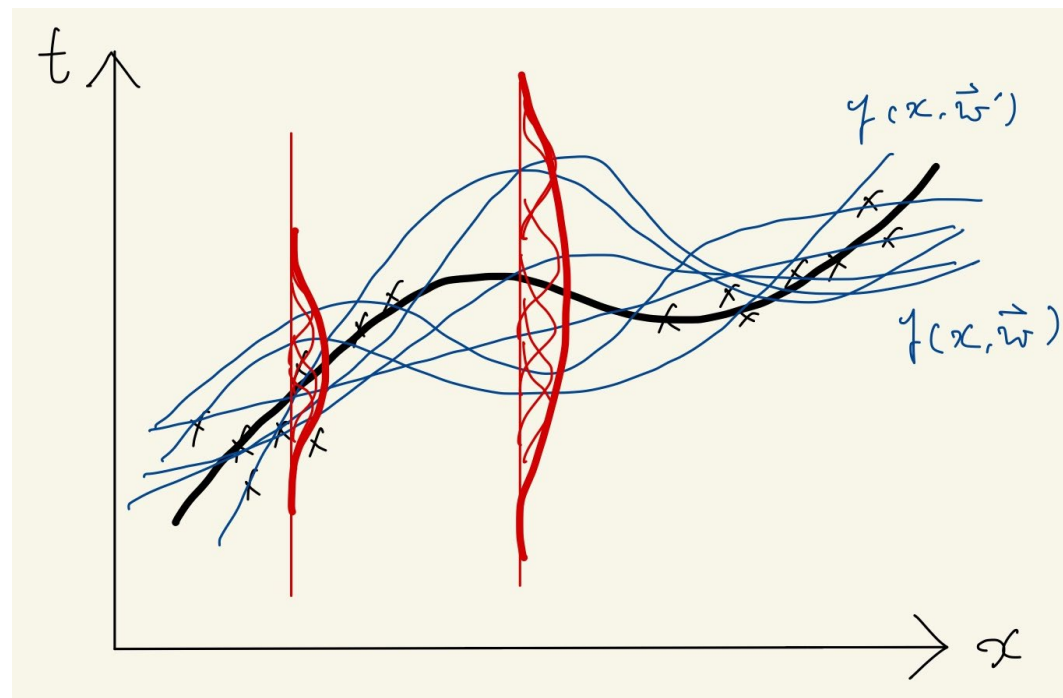
ベイズ推定の予測分布 (Predictive distribution)

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

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



必要な計算

予測分布 (Predictive distribution)

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

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

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

Bayes' theorem

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

Gaussianの積の積分もまたGaussianになる

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

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)

1次元のガウス分布

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \quad \begin{array}{l} \mu : \text{mean (平均)} \\ \sigma^2 : \text{variance (分散)} \end{array}$$

D次元のガウス分布

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

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

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

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

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

Covariance (共分散)

Covariance matrix Σ の性質

◆ Real symmetric (対称行列)

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

◆ Positive definite (正定値) 全ての固有値が正

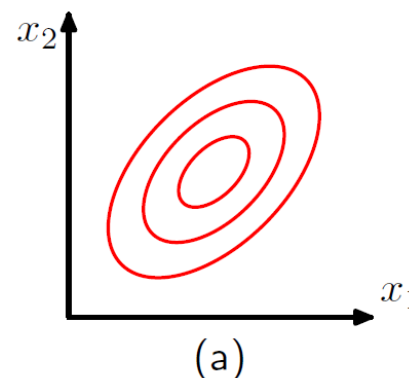
$$\Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad \lambda_i > 0$$

なぜなら

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

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

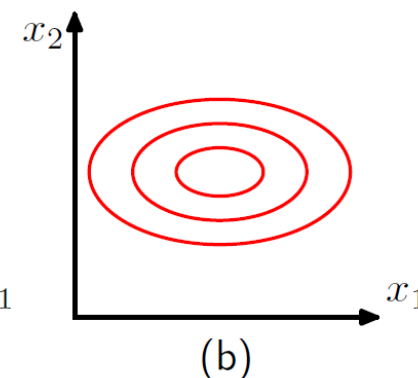
$$\Sigma_{12} > 0$$



正の相関

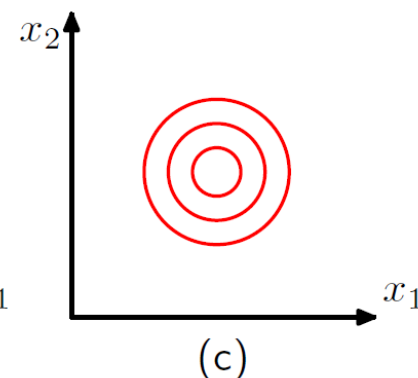
x_1 と x_2 が近い値
を取りやすい

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



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

PRML Fig. 2.8

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 $\boldsymbol{\Sigma}_{ab} = \boldsymbol{\Sigma}_{ba}^T$

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

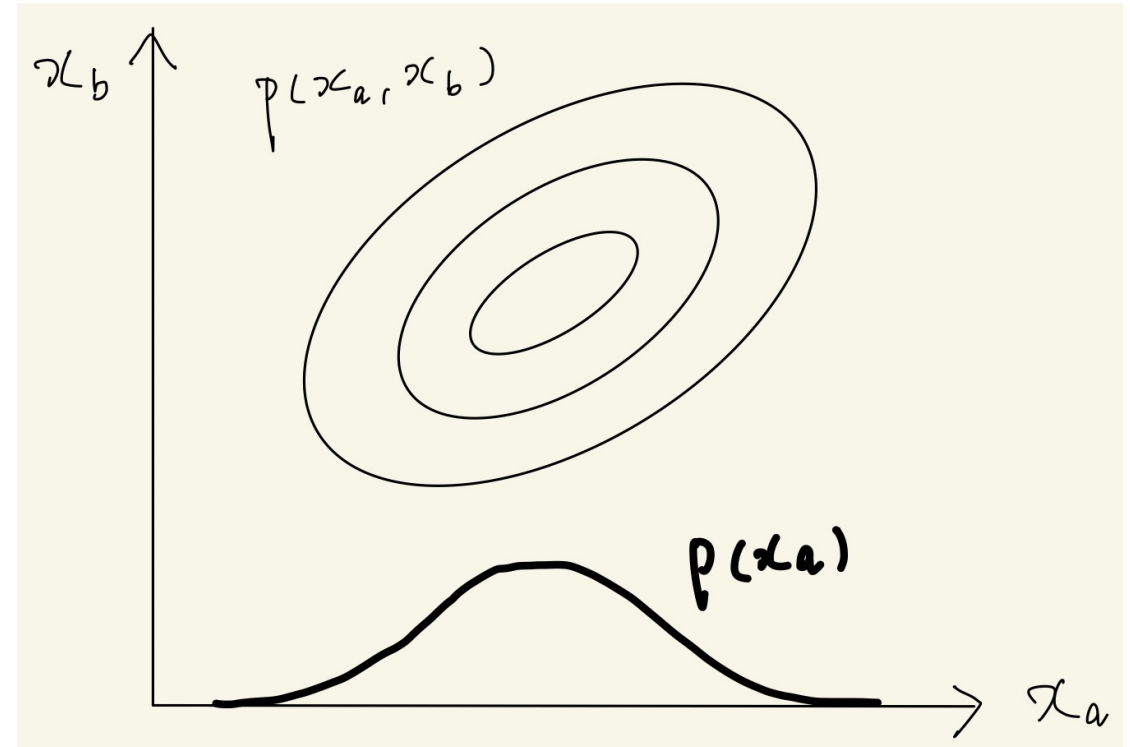
変数 $\mathbf{x}_b$ について周辺化（積分）する

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

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

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

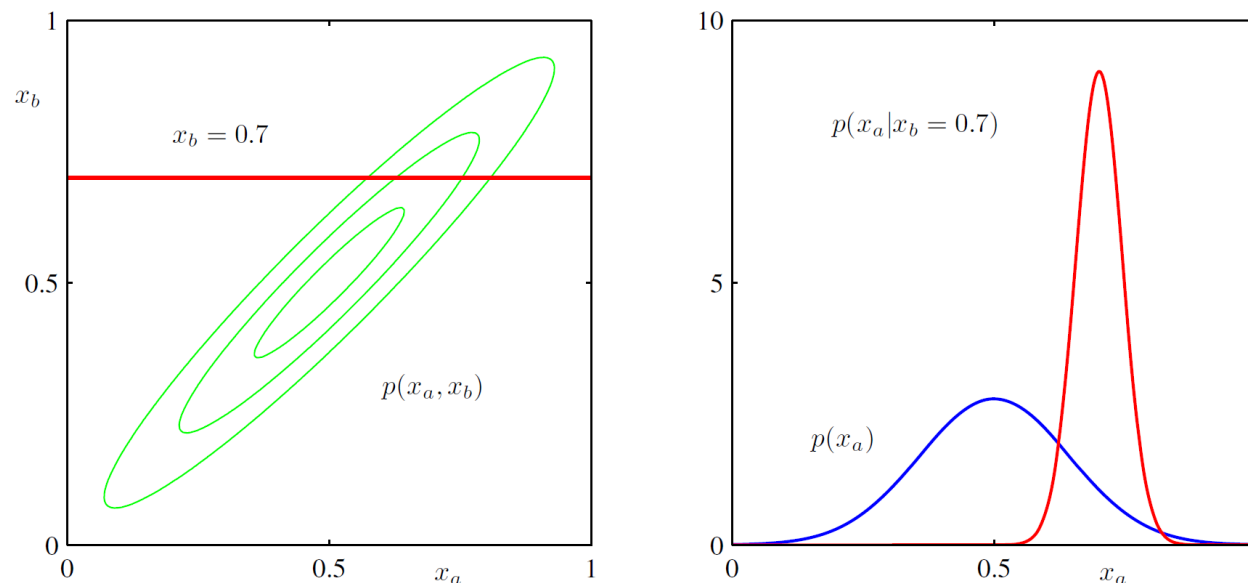
product rule

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Gaussian

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

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

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

Conditional distribution:

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

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

Marginal distribution:

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

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

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

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

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

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

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

where

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