



# システム制御理論と統計的機械学習

## 第7章：機械学習

---

加嶋 健司

October 3, 2025

京都大学情報学研究科

# 本章の流れ

## 7.1 回帰

未知の関数をその有限個の代入値から推定する。

## 7.2 特徴量抽出

制御におけるモデル低次元化などに直結する次元削減。

## 7.1 標本の生成と利用

確率変数の統計的性質を調べるために用いる標本の生成。

# 回歸

---

# 統計推論としての定式化 (1) 関数に対する統計推論

集合  $X$  上の実数値関数からなる集合  $\mathcal{F}$  を考え、推定対象の関数は  $f \in \text{rv}(\mathcal{F})$

- サイコロ  $f$  を投げると  $\varphi_f(f)$  に比例する実現しやすさで関数  $f \in \mathcal{F}$  が一つ定まる
- (何らかの  $x$  に対する) 代入値  $y = f(x)$  の観測結果のもとに、 $f$  の事後分布を求める

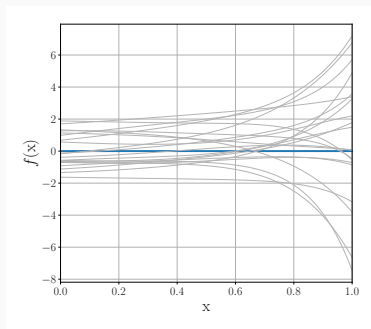


Figure 1: 事前分布  $f$

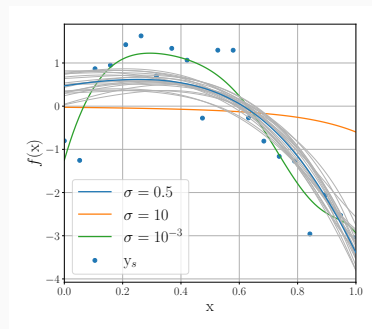


Figure 2: 事後分布  $f|_{y_1:\bar{s}}$

## 統計推論としての定式化 (2) 代入値のもとでの事後分布

- 代入する値  $x \in X$  を固定するごとに  $f(x) \in \text{rv}(\mathbb{R})$  が定まる
- 既知の評価点  $\{x_s\}_{s=1}^{\bar{s}}$  に対する加法的雑音下の代入値

$$y_{1:\bar{s}} := \begin{bmatrix} y_1 & \cdots & y_{\bar{s}} \end{bmatrix}^\top, \quad y_s := f(x_s) + w_s \quad (7.1)$$

を用いて  $f$  の推定を行う。ただし,  $f, w_1, w_2, \dots$  は独立であるとする。

- $f$  を潜在変数,  $y_{1:\bar{s}}$  を観測変数とする尤度関数は

$$\varphi(y_{1:\bar{s}}|f) = \prod_{s=1}^{\bar{s}} \varphi_{w_s}(y_s - f(x_s)) \quad (7.2)$$

- $y_{1:\bar{s}} = y_{1:\bar{s}}$  のもとでの  $f$  の事後分布は

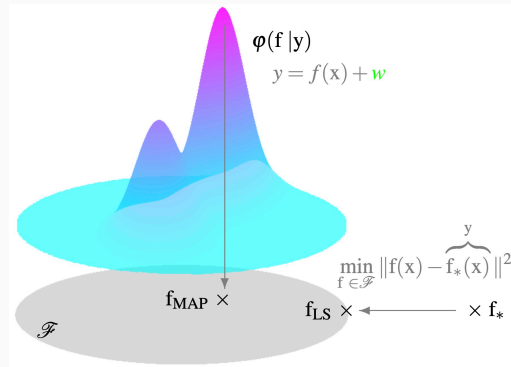
$$\varphi(f|y_{1:\bar{s}}) \propto \varphi(f) \prod_{s=1}^{\bar{s}} \varphi_{w_s}(y_s - f(x_s)) \quad (7.3)$$

## 統計推論としての定式化 (3) MAP 推定による回帰

したがって，観測結果  $y$  のもとでの  $\mathcal{F}$  上の事後分布

$$\varphi(f|y_{1:\bar{s}}) \propto \varphi(f) \prod_{s=1}^{\bar{s}} \varphi_{w_s}(y_s - f(x_s)) \quad (7.4)$$

を分布推定結果とし，この式を最大化する MAP 推定  $f_{\text{MAP}}$  などを点推定結果とする



**Figure 3:** 統計推論および最小二乗法による関数回帰

# 統計推論としての定式化 (4) 回帰モデル

## $\mathcal{F}$ の設定

- 関数形を仮定しないと、自由度が高いが、事前分布の設定や事後分布の計算が困難
- パラメータ  $p \in \Theta$  をもつ既知の関数に限定し

$$\mathcal{F} := \{f_p : p \in \Theta\} \quad (7.5)$$

確率変数  $p \in \text{rv}(\Theta)$  とすれば、 $\varphi_{f_p}(f_p) = \varphi(p)$  により  $f_p \in \text{rv}(\mathcal{F})$

- 少数のパラメータで表現される分布を用いる手法を**パラメトリック手法**、それ以外を**ノンパラメトリック手法**とよぶ.

- $p, w_1, w_2, \dots$  は独立ならば、対数尤度  $\log \varphi(y_{1:\bar{s}}|p) = \sum_{s=1}^{\bar{s}} \log \varphi_{w_s}(y_s - f_p(x_s))$  や、

事後分布  $\varphi(p|y_{1:\bar{s}}) \propto \varphi(p) \prod_{s=1}^{\bar{s}} \varphi_{w_s}(y_s - f_p(x_s))$  の最大化により推定がおこなえる

## 統計推論としての定式化 (5) 線形回帰

最重要例： $f_p$  が  $n_f$  個の既知の実数値関数  $\{\phi_i\}_{i=1}^{n_f}$  の線形和 ( $f_p$  が  $p$  に関して線形)

- 既知の関数  $\phi : X \rightarrow \mathbb{R}^{n_f}$  を用いて,

$$\mathcal{F}_\phi := \{f_p := p^\top \phi : p \in \mathbb{R}^{n_f}\}, \quad \phi := \begin{bmatrix} \phi_1 & \dots & \phi_{n_f} \end{bmatrix}^\top \quad (7.6)$$

- $X = \mathbb{R}^n$ ,  $\phi(x) = x$  であれば  $\mathcal{F}_\phi$  は線形関数からなる集合 (線形回帰)
- 関数の数  $n_f$  を **特徴次元**,  $\mathbb{R}^{n_f}$  を **特徴空間** (feature space), 特徴空間の要素を **特徴ベクトル** (feature vector), 写像  $\phi$  を **特徴写像** (feature mapping) とよぶ

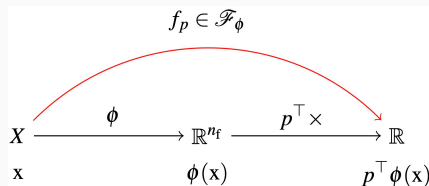


Figure 4: 線形パラメトリック回帰



# 統計推論としての定式化 (6) 多項式回帰

## 例 7.1.2 – 多項式回帰

左図は、 $X = [0, 1]$  および  $\phi(x) = [1, x, \dots, x^9]^\top$ ,  $p \sim \mathcal{N}(0, I)$  とした事前分布の標本関数とモード関数である．一方で、観測値  $\{y_s\}$  が与えられる場合に、 $w_s \sim \mathcal{N}(0, \sigma^2)$ ,  $\sigma = 0.5$  とした場合の事後分布の標本関数と MAP 推定関数を右図に示している．

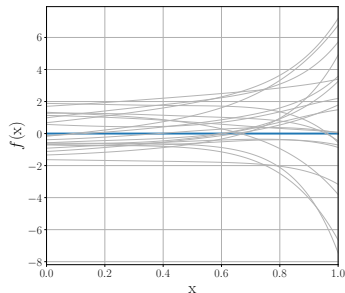


Figure 5: 事前分布  $f$

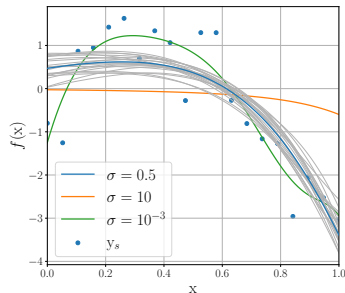


Figure 6: 事後分布  $f|_{y^{1:\bar{s}}}$

# 最小二乗法と正則化 (1) 最小二乗法

確率を用いない定式化

- 推定対象の関数  $f_* : X \rightarrow \mathbb{R}$  を  
 $\mathcal{F}_\phi := \{f_p := p^\top \phi : p \in \mathbb{R}^{n_f}\}$  の要素で近似
- 与えられた（雑音を含まない）代入値  
 $y_s := f_*(x_s)$  に対する二乗誤差を最小化

$$L_{LS}(p) := \sum_{s=1}^{\bar{s}} (y_s - f_p(x_s))^2$$

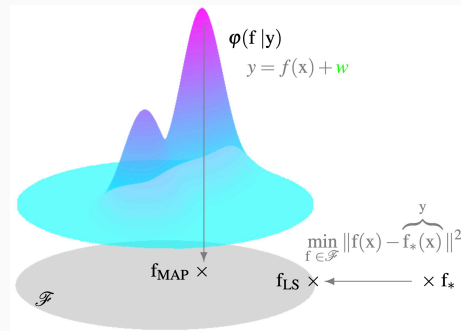


Figure 7: 統計推論および最小二乗法

## 最小二乗法と正則化 (2) 最小二乗法による線形回帰

- $\phi(\mathbf{x}_{1:\bar{s}}) := [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_{\bar{s}})] \in \mathbb{R}^{n_f \times \bar{s}}$  を用いて

$$\begin{aligned} L_{\text{LS}}(\mathbf{p}) &:= \sum_{s=1}^{\bar{s}} (y_s - \mathbf{p}^\top \phi(\mathbf{x}_s))^2 \\ &= \mathbf{p}^\top \phi(\mathbf{x}_{1:\bar{s}}) \phi(\mathbf{x}_{1:\bar{s}})^\top \mathbf{p} - 2\mathbf{p}^\top \phi(\mathbf{x}_{1:\bar{s}}) \mathbf{y}_{1:\bar{s}} + \text{定数} \end{aligned}$$

- 最小化解は**正規方程式**の解

$$\frac{1}{2} \nabla_{\mathbf{p}} L_{\text{LS}}(\mathbf{p}) = \phi(\mathbf{x}_{1:\bar{s}}) \phi(\mathbf{x}_{1:\bar{s}})^\top \mathbf{p} - \phi(\mathbf{x}_{1:\bar{s}}) \mathbf{y}_{1:\bar{s}} = 0 \quad (7.7)$$

- データが十分でない ( $\phi(\mathbf{x}_{1:\bar{s}}) \phi(\mathbf{x}_{1:\bar{s}})^\top$  が特異に近い) 場合、解が非常に大きくなる
  - データ数  $\bar{s}$  が特徴次元  $n_f$  より少ない場合は、解は非有界
  - 固有値が 0 に近いとデータ  $\mathbf{y}$  の変動に過度に鋭敏

$$(\phi(\mathbf{x}_{1:\bar{s}})^\top)^\dagger \mathbf{y}_{1:\bar{s}} = (\phi(\mathbf{x}_{1:\bar{s}}) \phi(\mathbf{x}_{1:\bar{s}})^\top)^{-1} \phi(\mathbf{x}_{1:\bar{s}}) \mathbf{y}_{1:\bar{s}} \quad (7.8)$$

- **正則化 (regularization)** : 評価関数に  $\mathbf{p}$  のノルムなど適切な項を追加する逆問題解析における常套手段

## 最小二乗法と正則化 (3) 統計推論との関係

### 定理 7.1.4 – 事前分布と正則化

$$w_s \sim \mathcal{N}(0, \sigma^2)$$

- 最尤推定は  $L_{LS}$  を最小化する最小二乗法
  - $p \sim \mathcal{N}(0, I)$  のとき, MAP 推定は  $L_{\text{Ridge}}(p) := L_{LS}(p) + \sigma^2 \|p\|^2$  を最小化するリッジ回帰
  - $\varphi(p) = \prod_i \text{Lap}(p_i | 0, 1)$  のとき, MAP 推定は  $L_{\text{Lasso}}(p) := L_{LS}(p) + \sigma^2 \|p\|_1$  を最小化する **Lasso** (least absolute shrinkage and selection operator)
- 
- 正則化と事前分布, 2-ノルムによる近似誤差評価と観測誤差の正規分布性が対応
  - 1-ノルム, 2-ノルムの両方を含む問題を**エラスティックネット** (elastic net)
  - リッジ回帰の唯一の最小化解は  $(\phi(x_{1:\bar{s}})\phi(x_{1:\bar{s}})^\top + \sigma^2 I)^{-1} \phi(x_{1:\bar{s}})y_{1:\bar{s}}$

# 最小二乗法と正則化 (4) 正則化の影響の例

## 例 7.1.5 – 正則化と事前分布

$X = [0, 1]$ ,  $\phi(x) = [1, x, \dots, x^9]^\top$  (多項式回帰)

- 正則化がない場合, 過度に振動する (過学習)
- $\sigma = 0.1$  としたリッジ回帰, Lasso の結果はともに滑らか

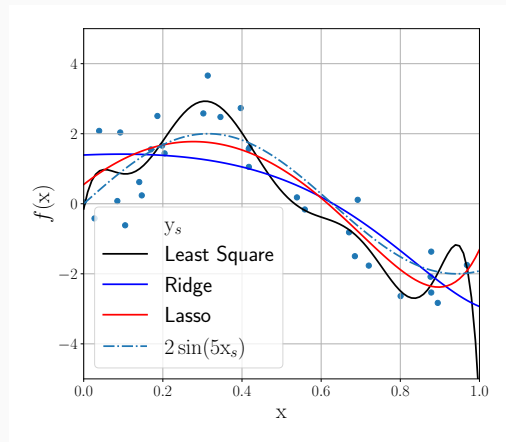
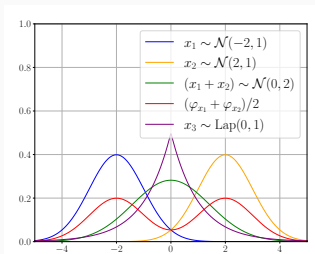


Figure 9: 回帰結果

# 最小二乗法と正則化 (5) 正則化の影響の例

## 例 7.1.5 – 正則化と事前分布 (つづき)

- 事前分布と最適解の関係（正則化がない場合の最大の係数は約  $4.5 \times 10^5$ ）
- Lasso に見られるような解の要素の多くが 0 となる性質は**スパース性**とよばれ、可解釈性向上などに用いられる。

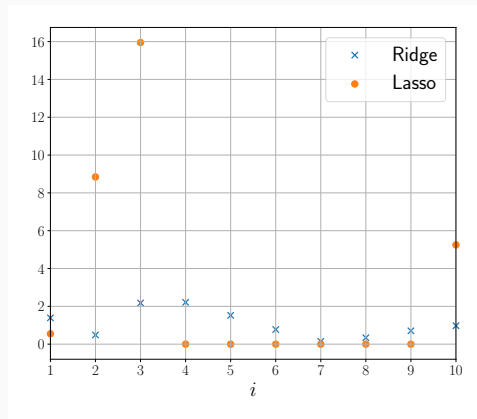
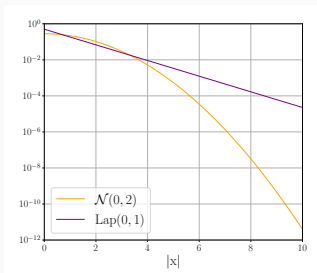


Figure 11: 回帰結果

## 最小二乗法と正則化 (6) カーネル行列

特徴次元  $n_f \gg$  データ数  $\bar{s}$  を想定した, 特徴写像  $\phi$  とリッジ回帰問題の構造

### 定理 7.1.8 – カーネルリッジ回帰

カーネル行列とよばれる

$$K_{\phi}(x_{1:\bar{s}}) := \begin{bmatrix} \phi(x_1)^{\top} \phi(x_1) & \phi(x_1)^{\top} \phi(x_2) & \cdots & \phi(x_1)^{\top} \phi(x_{\bar{s}}) \\ \phi(x_2)^{\top} \phi(x_1) & \phi(x_2)^{\top} \phi(x_2) & \cdots & \phi(x_2)^{\top} \phi(x_{\bar{s}}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(x_{\bar{s}})^{\top} \phi(x_1) & \phi(x_{\bar{s}})^{\top} \phi(x_2) & \cdots & \phi(x_{\bar{s}})^{\top} \phi(x_{\bar{s}}) \end{bmatrix} \quad (7.9)$$

を用いて,  $L_{\text{Ridge}}(p) := \sum_{s=1}^{\bar{s}} (y_s - p^{\top} \phi(x_s))^2 + \sigma^2 \|p\|^2$  を最小化する  $p$  は,

$$\phi(x_{1:\bar{s}})(K_{\phi}(x_{1:\bar{s}}) + \sigma^2 I)^{-1} y_{1:\bar{s}} \in \mathbb{R}^{n_f} \quad (7.10)$$

# 最小二乗法と正則化 (7) カーネルトリック

- $\phi(x), \phi(y)$  を計算し内積を求める計算量が大きくても、**カーネル関数**とよばれる

$$k(x, y) := \phi(x)^\top \phi(y), x, y \in X \quad (7.11)$$

が容易に計算できれば、カーネル行列の計算量が削減できる (**カーネルトリック**)

- 多項式カーネル:  $\phi(x) := \begin{bmatrix} 1 & \sqrt{2}x_1 & \sqrt{2}x_2 & \sqrt{2}x_1x_2 & x_1^2 & x_2^2 \end{bmatrix}^\top : \mathbb{R}^2 \rightarrow \mathbb{R}^6$

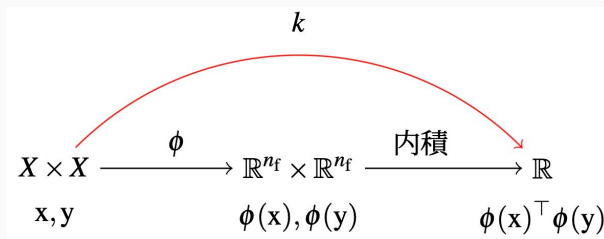


Figure 12: カーネル関数



## 最小二乗法と正則化 (8) リプレゼンター定理

### 注意 7.1.10 – リプレゼンター定理

最適なパラメータは必ず  $p = \phi(x_{1:\bar{s}})\rho$ ,  $\rho \in \mathbb{R}^{\bar{s}}$  と表現できる

特徴次元  $n_f$  がいかに大きくても，データの評価点  $\{x_s\}$  により定まる  $\bar{s}$  次元の特徴写像  $\phi'(\cdot) := \phi(x_{1:\bar{s}})^\top \phi(\cdot) : X \rightarrow \mathbb{R}^{\bar{s}}$  により最適な回帰関数は与えられ，

$$\|y_{1:\bar{s}} - p^\top \phi(x_{1:\bar{s}})\|^2 + \sigma^2 \|p\|^2 = \|y_{1:\bar{s}} - \rho^\top \phi'(x_{1:\bar{s}})\|^2 + \sigma^2 \rho^\top K_\phi(x_{1:\bar{s}}) \rho \quad (7.12)$$

## 特徴量抽出

---

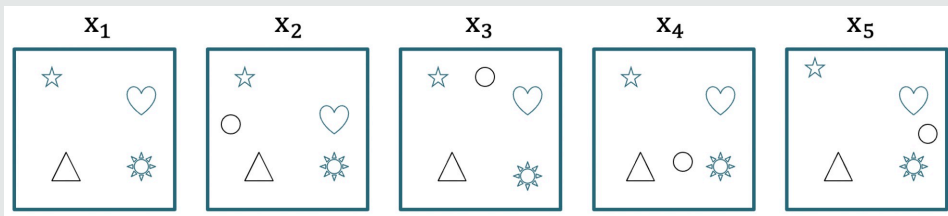
# 特徴量

与えられたデータ集合に対して、その分布の傾向を特徴づける量を抽出することを、機械学習分野では一般に**教師なし学習**（unsupervised learning）とよぶ。

- モデル低次元化などに直結する次元削減に焦点をあてる  
（クラスタリング，分類などは割愛）

## 例 7.2.1 – 特徴量

「星の位置」よりも「丸の位置」を各データの特徴量とすべき？



# 主成分分析 (1) 標本分散行列

データ集合  $\{x_s\}_{s=1}^{\bar{s}}$ ,  $x_s \in \mathbb{R}^d$  に対して,

**標本平均** 
$$\text{Ave}[\{x_s\}_{s=1}^{\bar{s}}] := \frac{1}{\bar{s}} \sum_{s=1}^{\bar{s}} x_s \in \mathbb{R}^d, \quad (7.13)$$

**標本分散行列** 
$$\text{Var}[\{x_s\}_{s=1}^{\bar{s}}] := \frac{1}{\bar{s} - 1} \sum_{s=1}^{\bar{s}} (x_s - \text{Ave}[x_{1:\bar{s}}]) (x_s - \text{Ave}[x_{1:\bar{s}}])^\top \in \mathbb{R}^{d \times d} \quad (7.14)$$

データ集合  $\{x_s\}_{s=1}^{\bar{s}} \subset X$  が与えられたとき, 実数値関数  $f$  に対する標本分散

$$\text{Var}[\{f(x_s)\}_{s=1}^{\bar{s}}] \in \mathbb{R} \quad (7.15)$$

が小さいとき, 各データ  $x_s$  の代わりに  $f(x_s)$  に注目してもデータ間の差異に関する情報は得られない (例えば,  $f$  が前ページの例の「星の位置」).

## 主成分分析 (2) 固有値問題への帰着

特徴写像を用いた関数  $f_p := p^\top \phi$  を対象とし,  $\|p\|^2 \leq 1$  により定数倍の自由度を制限

### 定理 7.2.2 – 主成分分析

$\{x_s\}_{s=1}^{\bar{s}} \subset X$  および特徴写像  $\phi : X \rightarrow \mathbb{R}_f^n$  が与えられたとき,  $f_p := p^\top \phi$  の標本分散

$$\text{Var}[\{f_p(x_s)\}_{s=1}^{\bar{s}}] \quad \text{subject to } \|p\|^2 \leq 1 \quad (7.16)$$

の最大値  $\lambda$  および最大化する  $p \in \mathbb{R}_f^n$  は

$$\text{Var}[\{\phi(x_s)\}_{s=1}^{\bar{s}}] \in \mathbb{R}^{n_f \times n_f} \quad (7.17)$$

の最大固有値および対応する単位固有ベクトルである.

- 平均 0 と平行移動 (中心化) された特徴写像  $\tilde{\phi}(x_s) := \phi(x_s) - \text{Ave}[\{\phi(x_s)\}_s]$
- $\tilde{\phi}(x_{1:\bar{s}}) := [\tilde{\phi}(x_1), \dots, \tilde{\phi}(x_{\bar{s}})] \in \mathbb{R}^{n_f \times \bar{s}}$  を用いて,  $\frac{1}{\bar{s}-1} \tilde{\phi}(x_{1:\bar{s}}) \tilde{\phi}(x_{1:\bar{s}})^\top \in \mathbb{R}^{n_f \times n_f}$

## 主成分分析 (3) 固有値分解による効率的計算

カーネルトリックの場合と同様の別表現

### 補題 7.2.3 – 標本分散行列の固有値分解

任意の行列  $M$  に対して、 $M^\top M$  の 0 でない固有値および対応する固有ベクトルを  $\lambda_i, \rho_i, i = 1, 2, \dots$ , とする. このとき、 $MM^\top$  の 0 でないすべての固有値・固有ベクトルは  $\lambda_i$  および  $p_i := M\rho_i$  により与えられる.

- $\tilde{\phi}(x_{1:\bar{s}})\tilde{\phi}(x_{1:\bar{s}})^\top \in \mathbb{R}^{n_f \times n_f}$  の固有値と固有ベクトルは、 $\tilde{\phi}(x_{1:\bar{s}})^\top \tilde{\phi}(x_{1:\bar{s}}) \in \mathbb{R}^{\bar{s} \times \bar{s}}$  の非零固有値と対応する固有ベクトル  $\rho$  を求め、 $p := \tilde{\phi}(x_{1:\bar{s}})\rho$  とすることで求められる.
- $k(x_s, x_{s'}) := \phi(x_s)^\top \phi(x_{s'})$  さえ計算できれば、 $\phi(x_s)$  の計算は不要

# 次元削減 (1) 固有値分解

## 固有値分解

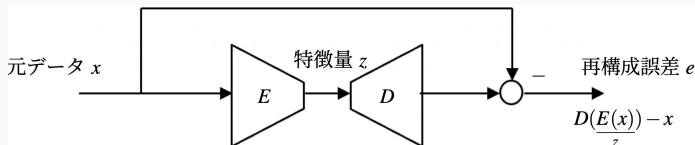
$$\text{Var}[x_{1:\bar{s}}] = \sum_{i=1}^n \lambda_i \xi_i \xi_i^\top, \|\xi_i\| = 1, \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

と固有値分解されるとき,  $\sum_s \|p^\top (x_s - \text{Ave}[x_{1:\bar{s}}])\|^2$  は  $p = \xi_1$  により最大化される.

これは各データ  $x \in \mathbb{R}^n$  をスカラー値  $\xi_1^\top (x - \text{Ave}[x_{1:\bar{s}}])$  に情報を圧縮

- **符号化器 (encoder)**  $E : \mathbb{R}^n \rightarrow \mathbb{R}^m$  により  $n$  次元データを  $m$  次元データに圧縮し,  
**復号化器 (decoder)**  $D : \mathbb{R}^m \rightarrow \mathbb{R}^n$  により復元する

$$\frac{1}{\bar{s}} \sum_{s=1}^{\bar{s}} \|D(E(x_s)) - x_s\|^2$$



## 次元削減 (2) 固有値分解による次元削減

### 定理 7.2.4 – 固有値分解による次元削減

$\{x_s\}_{s=1}^{\bar{s}}, x_s \in \mathbb{R}^n$  の標本共分散行列  $\text{Var}[x_{1:\bar{s}}]$  が

$$\text{Var}[x_{1:\bar{s}}] = \sum_{i=1}^n \lambda_i \xi_i \xi_i^\top, \|\xi_i\| = 1, \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

と分解できるとき、復元誤差を最小化するアファインな関数の組  $(E, D)$  の一つは、

$$E(x) := \Xi_m^\top (x - \text{Ave}[x_{1:\bar{s}}]), D(z) := \Xi_m z + \text{Ave}[x_{1:\bar{s}}] \quad (7.18)$$

で与えられる。ただし、 $\Xi_m := \begin{bmatrix} \xi_1 & \dots & \xi_m \end{bmatrix}$  である。

- 第  $m$  固有ベクトルまでで張られる部分空間に射影することが最適
- ニューラルネットワークによる符号化器・復号化器の組は**自己符号化器** (autoencoder) とよばれ、非線形の次元削減の実用的手法



# 標本の生成と利用

---

## 標本生成 (1) 逆関数法

- 一様分布や正規分布をはじめ、代表的な確率密度関数に対しては効率的な乱数生成法が知られており、さまざまな計算パッケージで利用できる。
- 任意に与えられた確率密度関数に対して、その乱数生成は必ずしも容易ではない。

### 定理 7.3.1 – 逆関数法

実数上の確率密度関数  $\bar{\varphi}$  の累積分布関数

$$\bar{F}(x) := \int_{-\infty}^x \bar{\varphi}(x') dx' = \mathbb{P}_{\bar{x} \sim \bar{\varphi}}(\bar{x} \leq x) \quad (7.19)$$

に対して、 $u \sim \text{Uni}([0, 1])$  に対して、 $x := \bar{F}^{-1}(u)$  の確率密度関数は  $\bar{\varphi}$  である。

- 累積分布関数の逆関数が具体的に計算可能であれば、一様分布にしたがう乱数の生成により所望の分布をもつ乱数が生成できる。

## 標本生成 (2) 棄却法

一方、つぎの**棄却法**はより一般の分布に適用できる。

### 定理 7.3.2 – 棄却法

実数上の確率密度関数  $\bar{\varphi}$  に対して、 $\bar{\varphi}(x) < c\varphi'(x)$ ,  $\forall x$  を満たす任意の定数  $c > 1$  および確率密度関数  $\varphi'$  をとる。このとき、

1.  $\varphi'$  にしたがう乱数  $y$  と  $\text{Uni}([0, 1])$  にしたがう乱数  $u$  を生成する、
2.  $\bar{\varphi}(y) \geq u c \varphi'(y)$  ならば  $x := y$  として終了し、そうでなければ 1. に戻る

とする。このとき  $x$  は  $\bar{\varphi}$  にしたがう乱数となる。

- **提案分布**とよばれる  $\varphi'$  は容易に乱数を生成できる確率密度関数
- $\varphi'$  は  $\bar{\varphi}$  とはまったく異なってもよいが、確率密度関数の規格化条件より  $c$  が非常に大きい値となるため、2. で棄却される確率が大きくなり生成効率が悪くなる。

# モンテカルロ法 (1) フィッシャー情報行列

与えられた分布の統計的性質を乱数生成を用いて調べる手法を総じて**モンテカルロ法** (Monte Carlo method) とよぶ。

## 定理 7.3.3 – 独立な観測に対するフィッシャー情報行列

確率変数  $x \in \text{rv}(X)$  に対して,  $\{y_s\}_{s=1}^{\bar{s}}$  は  $x$  のもとで条件付き独立であるとする。このとき,  $y_{1:\bar{s}}$  から  $x$  を推定する際のフィッシャー情報行列は

$$\text{FIM}_{y_{1:\bar{s}}|x=x} = \sum_{s=1}^{\bar{s}} \text{FIM}_{y_s|x=x}, \forall x \in X \quad (7.20)$$

- $\{y_s\}$  が条件付き独立同分布であれば, フィッシャー情報行列は試行回数に比例して増大し, 各回で得られる情報量が次第に減少するといったことは起こらない。

# モンテカルロ法 (2) 大数の法則・中心極限定理

乱数の個数が十分に多い場合の標本平均の性質

## 定理 7.3.4 – 大数の法則・中心極限定理

$x_s$  を平均  $\mu$ , 分散  $\sigma^2$  をもつ独立同分布にしたがう確率変数とする. このとき, 確率変数  $S_k := \sum_{s=1}^k x_s$  に対して, 以下のことが成り立つ.

- $S_k/k$  は  $\mu$  に確率収束する (**大数の弱法則**)
- $S_k/k$  は  $\mu$  に概収束する (**大数の強法則**)
- $(\sqrt{k}/\sigma)(S_k/k - \mu)$  は  $\mathcal{N}(0, 1)$  に分布収束する (**中心極限定理**)
- 確率変数  $z$  および関数  $V$  に対して  $\mathbb{E}[V(z)]$  を計算する状況において,  $z$  と同分布をもつ  $z_s$  を用いて標本平均  $\frac{1}{k} \sum_{s=1}^k V(z_s)$  を考える
- 大数の法則より,  $\mathbb{E}[V(z)]$  に収束する. また, 十分に大きい標本数  $k$  に関してそれらの差は正規分布にしたがい, その分散は  $1/k$  のオーダーで小さくなる

# 生成モデル

確率変数  $x \in \text{rv}(X)$  によりモデル化したい現象の独立な標本  $\{x_s\}$

低計算量の標本生成と柔軟な条件付けが可能な  $x$  の数学的モデル（**生成モデル** (generative model) とよばれる）の構築

## Example

- ・ 標本のヒストグラムを作成し、その概形をとらえた関数を  $\varphi(x)$  とする  
 $\Rightarrow$  分布の形状をとらえること自体が困難、標本生成も高計算負荷
- ・ 自己符号化器  $E$  を設計し、復号化器  $D: Z \rightarrow X$  を用いて  $x = D(z)$   
 $\Rightarrow D$  への代入は容易であるが、 $z$  の標本生成には符号化器を通した標本  $\{E(x_s)\}_s$  の分布情報が必要

代表的な手法：変分自己符号化器（VAE），敵対的生成ネットワーク（GAN），拡散モデル

# 第7章

## 1. 回帰

統計推論としての定式化

最小二乗法と正則化

## 2. 特徴量抽出

主成分分析

次元削減

## 3. 標本の生成と利用

標本生成

モンテカルロ法

生成モデル