情報認識
「ノンパラメトリック法」

- **担当教員： 杉山　将（計算工学専攻）**
- **居室：** W8E-505
- **電子メール:** sugi@cs.titech.ac.jp

これまでの復習

- **最大事後確率則:**

$$x \in \omega_k : k = \arg\max_i p(\omega_i \mid x)$$

- **ベイズの定理より**

$$p(\omega_i \mid x) \propto \underbrace{p(x \mid \omega_i)}_{\text{条件付き確率}} \underbrace{p(\omega_i)}_{\text{事前確率}}$$

- **事前確率は各カテゴリの標本の割合で推定**

$$\hat{p}(\omega_i) = n_i / n$$

- **どうやって条件付き確率を推定するか？(簡単のため，以下では条件付きでない確率 $p(x)$ を $\{x_i\}_{i=1}^n$ から推定する問題を考える)**

条件付き確率の推定1

- **パラメトリックモデル**を幾つか用意 $\{q_j(x;\theta)\}_j$
- **最尤推定法**により，パラメータを推定（手持ちのデータが最もよく現れるようにパラメータを決定）

$$\boxed{\hat{p}_j(x) = q_j(x;\hat{\theta}_{ML}^{(j)})}$$

$$\hat{\theta}_{ML}^{(j)} = \arg\max_{\theta\in\Theta_j} L_j(\theta)$$

$$L_j(\theta) = \prod_{i=1}^{n} q_j(x_i;\theta)$$

- **赤池の情報量規準**によりモデルを選択

$$\boxed{\hat{p}(x) = \hat{p}_a(x)}$$

$$a = \arg\max_j AIC(j)$$

$$AIC(j) = -\sum_{i=1}^{n} \log \hat{p}_j(x_i) + \dim \Theta_j$$

条件付き確率の推定2

- **パラメトリックモデル**を幾つか用意 $\{q_j(x \mid \theta)\}_j$ (パラメータは確率変数として扱う)
- **ベイズ推定法**により確率密度関数を推定(パラメータの事後分布でモデルを平均)

$$\hat{p}_j(x) = p(x \mid x_1, x_2, \ldots, x_n)$$

$$= \int_{\Theta_j} q_j(x \mid \theta) p_j(\theta \mid x_1, x_2, \ldots, x_n) d\theta$$

$$p_j(\theta \mid x_1, x_2, \ldots, x_n) = \frac{\prod_{i=1}^{n} q_j(x_i \mid \theta) p(\theta; \beta)}{\int_{\Theta_j} \prod_{i=1}^{n} q_j(x_i \mid \theta') p(\theta'; \beta) d\theta'}$$

条件付き確率の推定2(続き)

- **ラプラス近似**：積分の近似計算法
- **最大事後確率推定法**：積分を最頻値一点で近似

$$\boxed{\hat{p}_j(x;\beta) = q_j(x;\hat{\theta}^{(j)}_{MAP},\beta)}$$

$$\hat{\theta}^{(j)}_{MAP} = \arg\max_{\theta\in\Theta_j}\left[L_j(\theta)p_j(\theta;\beta)\right]$$

$$L_j(\theta) = \prod_{i=1}^{n} q_j(x_i;\theta)$$

- **経験ベイズ法（周辺尤度最大化）**によりモデル，事前分布を決定

$$\boxed{\hat{p}(x) = \hat{p}_a(x;\beta_{EB})}$$

$$(a,\beta_{EB}) = \arg\max_{(j,\beta)} ML(j,\beta)$$

$$ML(j,\beta) = \int_{\Theta_j}\prod_{i=1}^{n} q_j(x_i;\theta)p(\theta;\beta)d\theta$$

これまでの復習

- 最大事後確率則： $p(\omega_i \mid x) \propto p(x \mid \omega_i) p(\omega_i)$
- 事前確率の推定： $\hat{p}(\omega_i)$
- 条件付き確率の推定： $\hat{p}(x \mid \omega_i)$
  - パラメトリック法
    - 最尤推定法
    - ベイズ推定法，最大事後確率法
  - ノンパラメトリック法

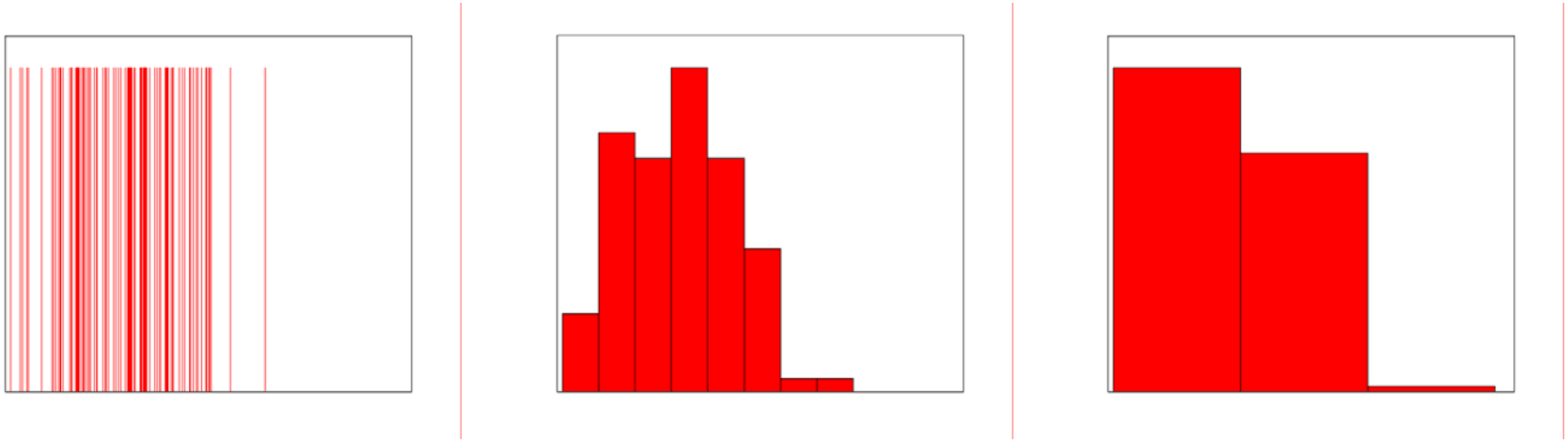
ノンパラメトリック法

- パラメトリック法：確率密度関数のモデルを用いて推定
- ノンパラメトリック法(non-parametric method)：モデルを用いないで，より直接的に確率密度関数を推定

ヒストグラム法

- **ヒストグラム法**(histogram method):**単純にヒストグラムを用いて確率密度関数を推定する方法**
  - **パターン空間を適当に分割する**
    (必ずしも同じ形に分割する必要はない).
  - **各分割内に入る訓練標本を数える.**
  - **積分が1になるように正規化する.**
- **非常に単純な方法なので便利であるが・・・**

ヒストグラム法の問題点



- 領域間で不連続
- 領域の分割の仕方を決めるのが難しい

もう少し工夫した方法が必要

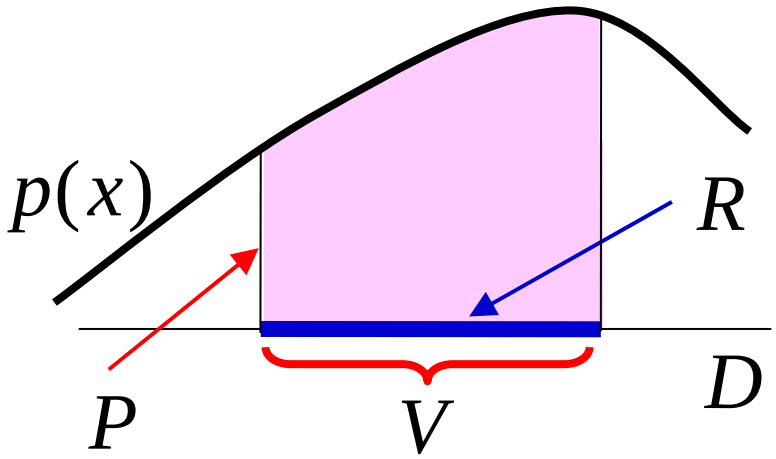
ノンパラメトリック法の表記

- R ：パターン空間 D 内のある領域(region)
- V ： R の体積(volume)

$$V = \int_R dx$$

- P ：あるパターン x が R に入る確率

$$P = \int_R p(x)dx$$



ノンパラメトリック法の基礎

- 推定したい確率密度関数 $p(x)$ が R 内でほぼ一定値を取るとき，確率 P は領域 R 内のある点 x を用いて

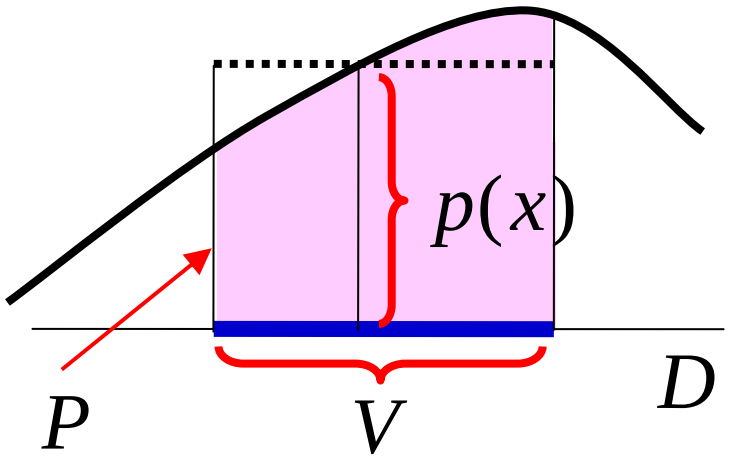
$$P \approx Vp(x)$$

- k ：n個の訓練標本のうち R に入っている個数

$$P \approx k/n$$

$$\Rightarrow \quad p(x) \approx \frac{k}{nV}$$

積分の長方形近似



領域の決め方

$$p(x) \approx \frac{k}{nV}$$

- 近似のよさは領域 R の決め方 (k,V) に依存.
- $p(x)$ が R 内でほぼ一定値をとる

  ⇒ R をできるだけを小さくする
- しかし R を小さくしすぎると, k が0か1しか取らなくなってしまい, $P \approx k/n$ の近似精度が悪い.

⇓

程よい大きさに R を決める必要がある

領域の決め方 (続き)

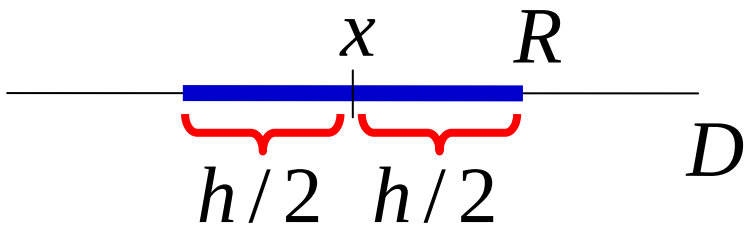
$$p(x) \approx \frac{k}{nV}$$

- 訓練標本を用いて領域 R を決める.
  - パーゼン窓法, 核密度推定法:
    V を固定したもとで k を標本から決定
  - 最近傍密度推定法:
    k を固定したもとで V を標本から決定

パーゼン窓関数

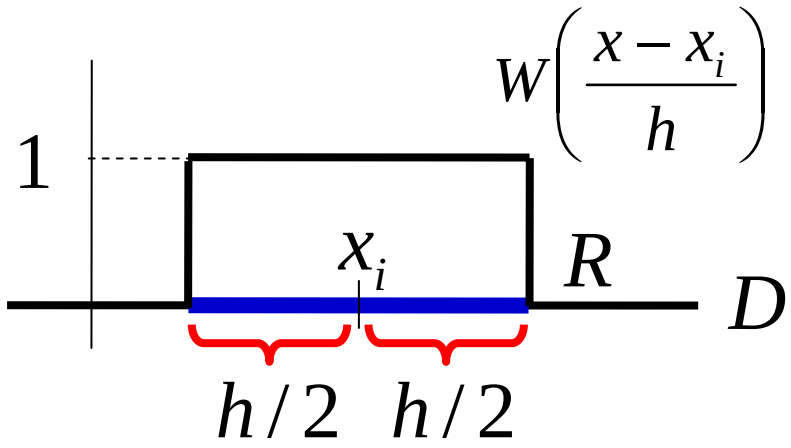
- 領域 R として，ある点 x を中心とする一辺の長さが h の**超立方体**(hypercube)を用いる．
  - 体積 $V = h^d$
  - R に入る標本数

$$k = \sum_{i=1}^{n} W\left(\frac{x - x_i}{h}\right)$$



$$W(x) = \begin{cases} 1 & \max|x^{(i)}| < \frac{1}{2} \\ 0 & \text{それ以外} \end{cases}$$

パーゼン窓関数
(Parzen window function)

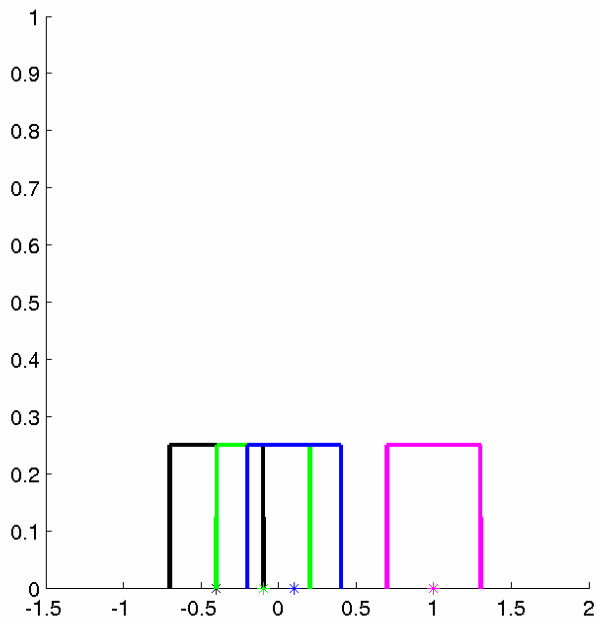


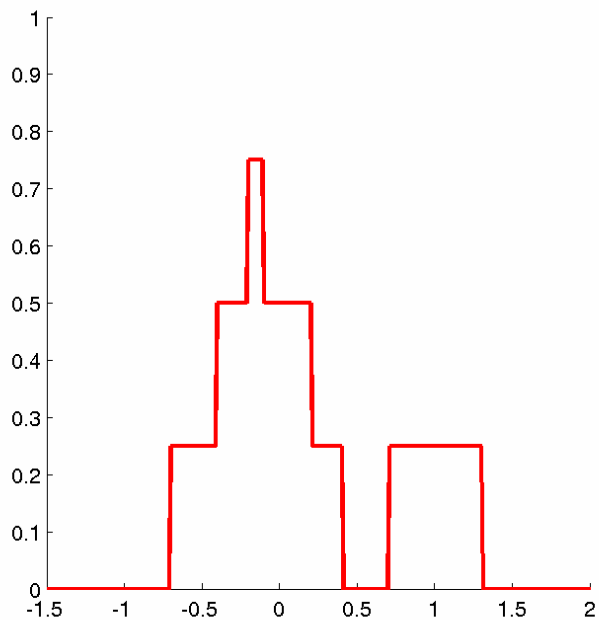
パーゼン窓法

■ **パーゼン窓法**(Parzen window method):

$$\hat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} W\left(\frac{x - x_i}{h} \right)$$

$$p(x) \approx \frac{k}{nV}$$





パーゼン窓法の特徴

- 与えられた標本から，パターン領域の分割を適応的に決定できるため，ヒストグラム法のようにあらかじめ領域の分割の仕方を決定しておく必要が無い．
- 領域間での不連続性は未解決．
- 領域の分割の仕方を決める必要は無いが，パーゼン窓関数の幅 h を適切に決める必要がある．

小レポート：計算機実験

- パーゼン窓法を実装し，適当なデータを用いて確率密度関数を推定せよ．
- データ標本数，真の確率分布，窓関数の幅などの条件を変化させたとき，どのように推定結果が変わるかを考察せよ．