

情報認識 「まとめ」

- 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

■ 識別関数のよさを測る規準

- 最大事後確率則
- 最小誤識別率則
- ベイズ決定則

■ 条件付き確率の推定

- パラメトリック法
 - 最尤推定法, EMアルゴリズム
 - ベイズ推定法, 最大事後確率推定法
- ノンパラメトリック法
 - カーネル密度推定法
 - 最近傍密度推定法

■ 最大事後確率則:

$$\arg \max_y p(y | x)$$

■ ベイズの定理より

$$p(y | x) \propto p(x | y) p(y)$$

条件付き確率 事前確率

■ 事前確率はそのカテゴリに含まれる標本の割合で推定

$$\hat{p}(y) = \frac{n_y}{n}$$

条件付き確率の推定1

281

- 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)

$$\hat{p}_j(x) = q_j(x; \hat{\theta}_{ML_j})$$

$$\hat{\theta}_{ML_j} = \arg \max_{\theta \in \Theta_j} L_j(\theta)$$

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

- 赤池の情報量規準によりモデルを選択

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

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

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

条件付き確率の推定2

282

- **パラメトリックモデル**を幾つか用意 $\{q_j(x|\theta)\}_j$
(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)

$$\hat{p}_j(x) = p(x | x_1, x_2, \dots, x_n)$$

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

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

条件付き確率の推定3

283

■ **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

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

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

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

■ **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

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

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

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

■ ノンパラメトリック法: パラメトリックモデルを用いない確率密度関数の推定法

- パーゼン窓法, カーネル密度推定法: 体積を固定し, 標本数をデータから決定する

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

- 最近傍密度推定法: 標本数を固定し, 体積をデータから決定する

$$\hat{p}(x) = \frac{k\Gamma(\frac{d}{2} + 1)}{n\pi^{\frac{d}{2}}h^d}$$

- カーネル密度推定法の式から, k-最近傍識別器が得られる. kの値は交差確認法により, 決定できる.

生成モデル vs. 識別モデル

285

$$p(y | x) \propto p(x, y) = p(x | y)p(y)$$

カテゴリの事後確率
(パターンの識別モデル)

データの同時確率
(データの生成モデル)

- 訓練標本は事前確率や条件付き確率に従って生成されているため、訓練標本からこれらの確率を推定する方法は**生成モデル(generative model)に基づく方法**とよばれている.
- 一方、生成モデルは直接推定せずに、決定境界を直接モデル化する方法もあり、これは**識別モデル(discriminative model)に基づく方法**とよばれている.

- **Vapnikの原理**: ある問題を解くとき, その問題よりも難しい問題を途中段階で解いてはならない
- 生成モデルが分かれば識別関数を構成できるが, 逆はできない.
- 生成モデルを求める問題は, 識別関数を求める問題よりも難しい?
- 識別関数を直接学習したほうがよい?

試験について

287

- 2月1日(月)10時40分～12時10分
- S222講義室
- 試験内容:
 - 専門用語の英単語(日本語→英語)
 - 自由記述問題(以下より2問選択)
 - 識別関数のよさを測る規準について
 - 最尤推定法について
 - ノンパラメトリック法について
 - モデル選択について
 - 自分が考えたパターン認識の例について
- 教科書, ノートの持ち込みは不可.