パターン認識と学習

岡山大学　竹内孔一

1

本日の内容

- 識別部の設計
  - 最尤法

2

4章全体

- 事例 x に対する分布を仮定
  - どう解析的に求まるか? 4.1, 4.2
  - 4.1 ノンパラメトリックとの比較
  - 4.2 最尤推定
  - どのようなアイデアを盛り込んだ拡張ができるか 4.3

3

確認

- 下記を最大にするとはどういう意味か考えよ

$$p(\mathcal{X};\phi)=\prod_{j=1}^{n}p(\mathbf{x}_j;\phi)$$

$$p(\mathcal{X}\mid A_i;\phi_i)=\prod_{j=1}^{n}p(\mathbf{x}_j\mid A_i;\phi_i)$$

- p.53, 54 の $\hat{\mathbf{m}}$, $\hat{\boldsymbol{\Sigma}}$ はどこから求まるか

4

最尤推定

- 以下の計算が最尤推定から得られることを証明する

$$P(s_i\mid s_{i-1})=\frac{C(s_{i-1},s_i)}{C(s_{i-1})}\qquad P(w_i\mid s_i)=\frac{C(s_i,w_i)}{C(s_i)}$$

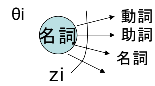
  - 最尤推定(maximum likelihood estimation)とは
    - 分布型Pに関する先験情報{θ1, …, θN}があるとき
      Y回の観測事象{x1,..,xY}に対して生起確率
      $$\prod_{k=1}^{Y}P(x_k;\theta_1,\cdots,\theta_N)$$
      またはその対数確率
      $$\sum_{k=1}^{Y}\log P(x_k;\theta_1,\cdots,\theta_N)$$
      を最大にする{θ1, …, θN}を推定する

5

最尤推定

- 推定対象の確率変数
  {θ1, …, θN}
- Y個の学習データ
- θi の出現回数が ni
- θi に関係する全事象の集合zi
- 事象の生起確率
  {P(θ1), …, P(θN)}→　これを最尤推定

P(助詞|名詞)の場合
θi は 名詞→助詞
ni は 名詞 から 助詞への遷移の回数



名詞→助詞に関連する
全事象の集合:
zi={名詞→動詞,名詞→名詞,..}

事象Nの出現回数がni回であるような標本集合の総確率値の最大化

$$L=\sum_{i=1}^{N}n_i\log P(\theta_i)$$

制約
$$\sum_{i\in Z_i}P(\theta_i)=1$$
Lagrange 未定乗数法

6

最尤推定

Lagrange 関数

$$F=\sum_{i=1}^{N} n_i \log P(\theta_i)-\lambda\left(\sum_{i\in Z_i} P(\theta_i)-1\right)$$

これをθi で偏微分する

$$\frac{\partial F}{\partial P(\theta_i)}=\frac{n_i}{P(\theta_i)}-\lambda=0 \quad \text{よって} \quad P(\theta_i)=\frac{n_i}{\lambda} \quad ①$$

制約式に代入して

$$1=\sum_{i\in Z_i} P(\theta_i)=\frac{1}{\lambda}\sum_{i\in Z_i} n_i \qquad \lambda=\sum_{i\in z_i} n_i$$

よってλを①に代入して

$$P(\theta_i)=\frac{n_i}{\sum_{i\in z_i} n_i}$$

参考：奈良先端大：
音声情報処理講義録
鹿野先生 1994年

7

ラグランジュ未定乗数法

- 制約の中で式を解く
  - M個の制約条件 $g_i(x)=0$ (i=1,2,..,M)
  - f(X) の極値をとりたい (極大or極小)

  ラグランジュ乗数 $\boldsymbol{\alpha}=[\alpha_1,..,\alpha_M]^t$ を使って

$$L(X,\boldsymbol{\alpha})=f(X)-\sum_{i=1}^{M}\alpha_i g_i(X)$$

$$\frac{\partial L}{\partial x_k}=0 \qquad \frac{\partial L}{\partial \alpha_i}=0$$

$(k=1,2,\ldots,N)$ $(i=1,2,\ldots,M)$ を解けばよい.

8