# パターン認識と学習

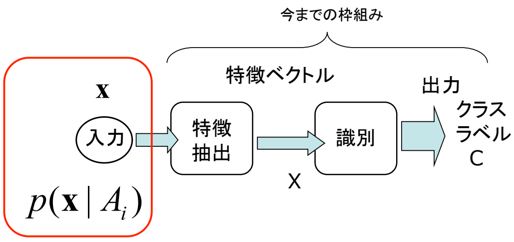
岡山大学大学院　竹内孔一

# 本日の内容

- 識別部の設計

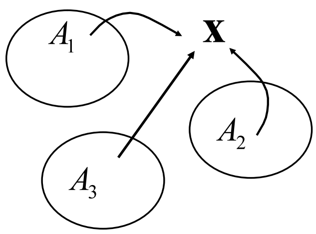
# 情報源の仮定

- 事例が統計的な分布に従う




前提: 事例 x は各クラスAiに対してある確率密度分布に従って生起していると仮定する

# 情報源の仮定




$P(A_i)$ 事前確率

$P(A_i \mid \mathbf{x})$ 事後確率

$p(\mathbf{x})$ xの生起確率

$p(\mathbf{x} \mid A_i)$ 各クラスからのxの生起分布

$$P(A_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid A_i)}{p(\mathbf{x})} P(A_i)$$

Bayes theorem

# 識別関数

- 事後確率を最大にするクラスを判別結果とする

$$\max_i \{P(A_i \mid \mathbf{x})\} = P(A_l \mid \mathbf{x})$$

未知の入力x はクラス Al に分類される

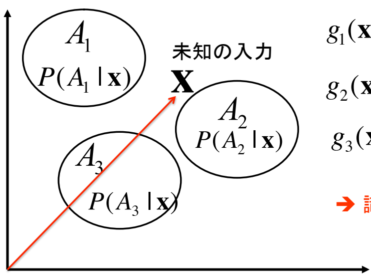
➔ つまり, $P(A_i \mid \mathbf{x})$ を識別関数と同じと見なす

$$g_i(\mathbf{x}) = \frac{p(\mathbf{x} \mid A_i)}{p(\mathbf{x})} P(A_i)$$

$$\stackrel{def}{=} p(\mathbf{x} \mid A_i) P(A_i)$$

$p(\mathbf{x})$はクラスAiを選択するときは固定なので省略

# 事後確率の最大化とは?



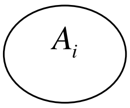

$$\left.\begin{array}{l} g_1(\mathbf{x}) = P(A_1 \mid \mathbf{x}) \\ g_2(\mathbf{x}) = P(A_2 \mid \mathbf{x}) \\ g_3(\mathbf{x}) = P(A_3 \mid \mathbf{x}) \end{array}\right\}$$ 最大値を判別結果とする

➔ 識別関数と同じこと

## 練習

- 各クラスAi内である特徴量xが生起する確率は正規分布であるとする
  - 平均ベクトル mi
  - 共分散行列 Σi

  とするとき，下記の図にその様子を書き込め
  (mi, Σiを使うこと)




•共分散行列 Σi=Σ0 とはどういう状況か説明せよ
•共分散行列 Σ0=I (単位行列) とはどういう状況か説明せよ

## 例題

- 入力が各クラスの正規分布で発生していると仮定したとき，NN法(最小距離識別法)はどういう特殊な条件であるか答えよ

## 識別部の設計

- 確認
  - 前提としてある事例 $\mathbf{x}_p$ はクラスAi からどのように生成されていると考えられるか
  - 共分散行列 (covariance matrix)はどういうものか説明せよ
  - マハラノビス距離 (Mahalanobis distance)はどのような特徴を測ったもの?

## 生成モデルと判別モデル

- 仮定と事後確率

判別モデル $p(A_i \mid \mathbf{x})$ 事後確率 -> 識別関数

生成モデル $p(\mathbf{x} \mid A_i)$ 仮説としておいた分布

やろうと思えば学習データから直接どちらも求まる

Bayes の定理を利用してp(x|Ai)を通してp(Ai|x)を求める方法を生成モデルと呼ぶ. Hidden Markov Modelなど.

直接学習データから p(Ai|x)を求める方法もある.事後確率を直接求めると分母が疎なので推定が(普通)よくないが, Conditional Random Field などでは exp()を分布として仮定して疎なデータに対して良い予測を与える．判別モデルと呼ばれる.

## パラメータ推定

- 下記の式の意味を答えなさい

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

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

- パラメトリックな学習とは何か?

## 最尤推定

- 全学習データXに対して下記の式が最大になるようにパラメータΦを学習する
  - そもそもなぜ最大にするのだろう? $p(X;\phi) = \prod_{j=1}^{n} p(\mathbf{x}_j;\phi)$

$p(X;\phi)$ はΦによってp(x)の形が決まり，全学習データXの確率値を書けた確率値(いつものパターン!)

↓

p(x;Φ)は未知だけど，実際にデータX={x1,x2,…xn}は出現した ➔ これが出てきたのだからデータXをp(x;Φ)はうまく説明して欲しい

↓

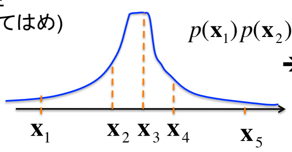
今，xは分布p(x;Φ)に依存して出力してると考えている.
ならば，p(x;Φ)からXが出力したと考えると，そのpで全データXに対して確率値p(X;Φ)を求めれば，最大の値になるはず

# 具体例

学習データ $X=\{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4,\mathbf{x}_5\}$
が下記のように分布しているとする

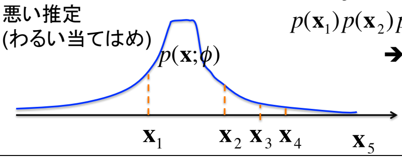
最尤推定→
密なところに大きな
確率値を割り当てる
ようにΦを決定

良い推定
(良い当てはめ)




$p(\mathbf{x}_1)p(\mathbf{x}_2)p(\mathbf{x}_3)p(\mathbf{x}_4)p(\mathbf{x}_5)$
→値が大きい

悪い推定
(わるい当てはめ)




$p(\mathbf{x}_1)p(\mathbf{x}_2)p(\mathbf{x}_3)p(\mathbf{x}_4)p(\mathbf{x}_5)$
→値が小さい

# 最尤推定(確率最大化→数を数える)

- 推定対象の確率変数
  {θ1, …, θN}
- N個の事象
  {N1,…, Nn}
- 事象 N の出現回数が ni
- 事象の生起確率
  {P(θ1), …, P(θN)} → これを最尤推定
  次の式を以下の制約のもとに最大化する

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

制約 $\sum_{i=1}^{N} P(\theta_i)=1$ Lanrange 未定乗数法

P(si|si-1)の場合
θi は 例えば si|si-1
Ni は si-1 から si の遷移
ni は si-1 から siの遷移の回数
C(si-1 → si )

# 最尤推定

Lagrange 関数

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

これをθi で偏微分する

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

制約式に代入して

$$1=\sum_{i=1}^{N} P(\theta_i)=-\frac{1}{\lambda}\sum_{i=1}^{N} n_i \qquad \lambda=-\sum_{i=1}^{N} n_i$$

よってλを①に代入して

$$P(\theta_i)=\frac{n_i}{\sum_{k=1}^{N} n_i}$$

参考:奈良先端大:
音声情報処理講義録
鹿野先生 1994年

# 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 $\boldsymbol{\alpha}=[\alpha_1,..,\alpha_M]'$ を使って

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

$$\frac{\partial L}{\partial x_k}=0 \quad (k=1,2,\ldots,N) \qquad \frac{\partial L}{\partial \alpha_i}=0 \quad (i=1,2,\ldots,M)$$

を解けばよい.