06T4029T 斉藤優

第3章 識別関数の良さを測る基準

3.1 訓練標本を用いた識別関数の学習
3.2 最大事後確率則
3.3 最小誤識別率則

識別関数

パターン認識器を構成　= 識別関数を構成

パターンXやカテゴリYの統計的性質を利用して、最適な識別関数を構成する

訓練標本

- 訓練標本

1. カテゴリが既知であるパターン x_i

$$\{(x_i, y_i)\}_{i=1}^{n}$$

2. カテゴリyに属するパターンの数 n_y

$$\sum_{y=1}^{c} n_y = n$$

主成分分析

- 主成分分析

1. 高次元のデータを低次元に圧縮

2. 2次元のデータなら決定境界を構成しやすい

3. 分散共分散行列の固有値と対応する固有ベクトルで2次元に圧縮

$$\begin{pmatrix} \phi_1^T x_i \\ \phi_2^T x_i \end{pmatrix}$$

決定境界

主成分分析により、2次元化されたデータの決定境界はどのようにすればよいのか？

データを正しく分類するのではなく、将来現れる未知のパターンを正しく分類すること、つまり凡化能力を獲得するように決定境界を定める.

最大事後確率則

・最大事後確率則

1. 識別関数の良さを測る規準

2. 事後確立が最大のカテゴリを選ぶ

$$\hat{y} = \arg\max_{y} p(y \mid x)$$

3. 条件付き確率と事前確率の積が最大のカテゴリにパターンxを分類

$$p(y \mid x) \propto p(x \mid y)\, p(y)$$

最小誤識別率(1)

- 最小事後確率則

1. 識別関数の良さを測る規準

2. パターンが誤って分類される確率を最小にするよう識別関数を決定

   - 誤識別率

$$p_e(y) = 1 - \int_{x \in D_y} p(x \mid y)\,dx$$

   - 全体の誤識別率

$$p_e = \sum_{y=1}^{c} p_e(y)\,p(y)dx$$

$$p_e = 1 - \sum_{y=1}^{c} \int_{x \in D_y} p(y \mid x)\,p(x)\,dx$$

最小誤識別率(2)

全体の誤識別率を p_e を最小にするには $\sum_{y=1}^{c} \int_{x \in D_y} p(y \mid x) p(x) dx$ を最大とすればよい

⇩

最大となるように決定領域 $\{D_y\}_{y=1}^{c}$ を決める.

⇩

誤識別率を最小にすることと事後確率を最大にすることは等価