

プロジェクト実習 III

パターン認識

- 第2週目 -

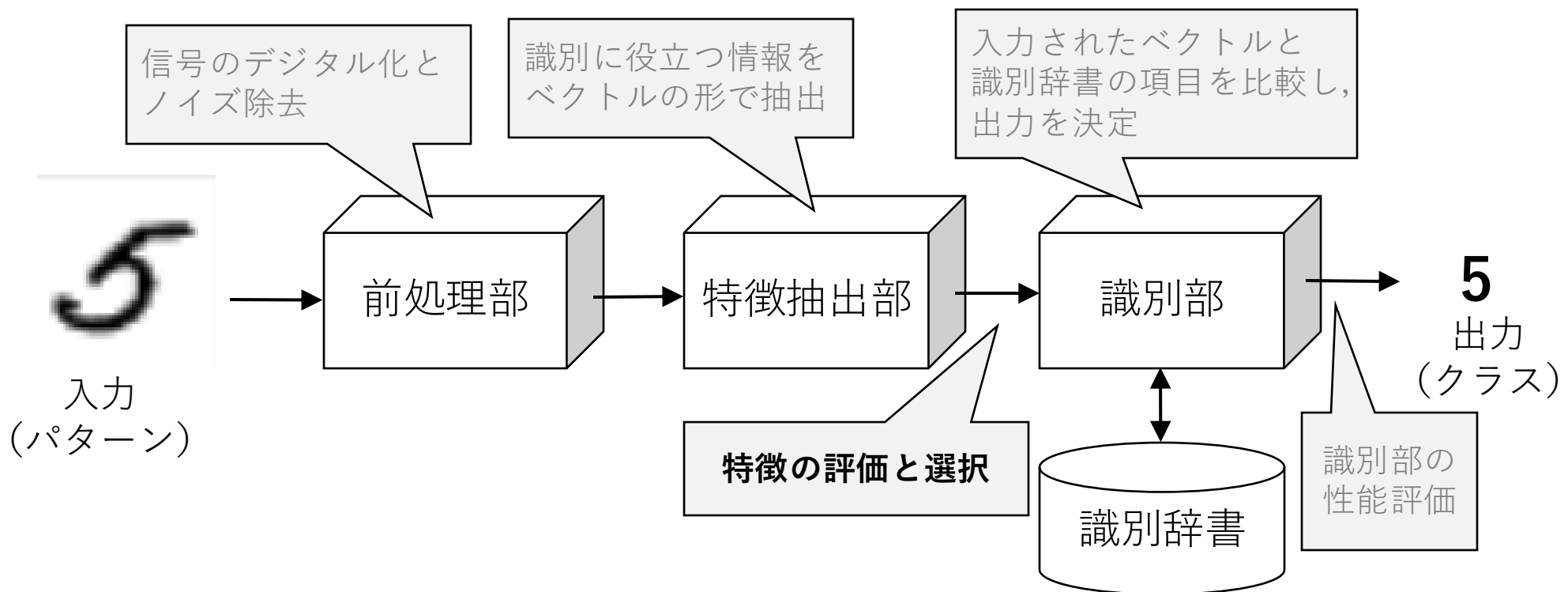
担当：崔恩瀨

パターン認識テーマ 4週間の計画

週	提出物	実験内容	テキスト
1		特徴抽出	1章
2		特徴の評価	2章
3	レポート(1,2週分)	数字識別	3章
4		識別性能の評価	4章
	レポート(3,4週分)		

- 提出期限：締切日の12:50
- コーディングはすべて Google Colaboratory で行う

第2週の実験



特徴抽出（第1週の実験内容）

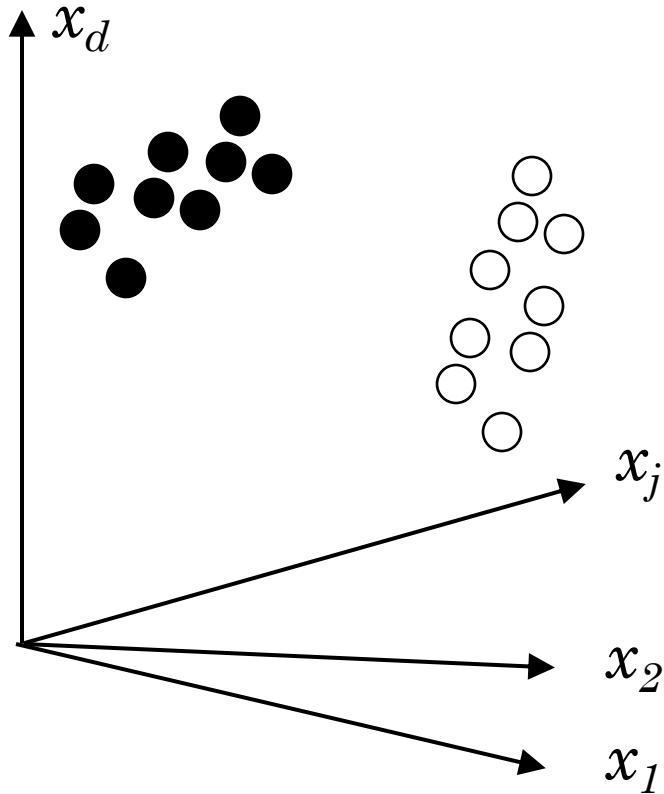
- 特徴抽出部の入出力
 - 入力：デジタル信号
 - 出力：パターンの特徴を表す d 次元ベクトル $\boldsymbol{x} = (x_1, x_2, \dots, x_d)^T$
- 抽出した特徴ベクトル
 - 各画像について縦方向・横方向の重心・分散・ゆがみ・扁平度
 - 標準化処理によって各特徴は平均0，分散1に変換済み

抽出した特徴の評価

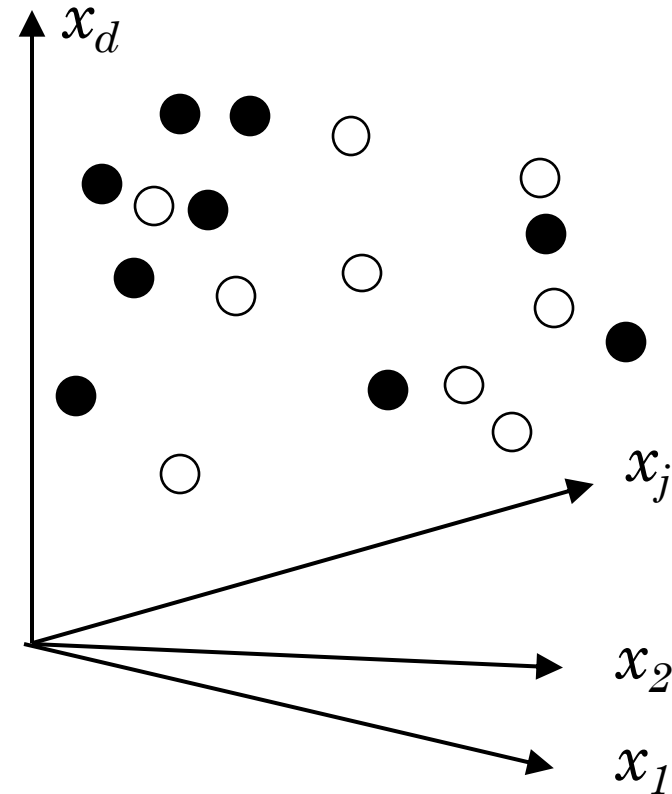
- 特徴空間における評価

適切な特徴であれば、同じクラスのデータは特徴空間上でクラスタとなる

識別に適した特徴



識別に適さない特徴



抽出した特徴の評価

- 特徴の評価法

- 高次元空間のデータは可視化できないので、定量的に特徴空間の良さを表したい
 - 定量評価：クラス内分散・クラス間分散比
- 定量的な評価法が完全なものでない場合、定性的な評価法と組み合わせる
 - 定性評価：2次元散布図のプロット

定量評価

- クラス内分散・クラス間分散比
 - 選択した特徴の評価法
 - 特徴空間の評価法でクラス毎のデータの広がり方を評価する尺度
- 同じクラスのデータはなるべく接近し，異なるクラスのデータはなるべく離れているものが高い値を取るようにする

定量評価

- クラス内分散

$$\sigma_W^2 = \frac{1}{n} \sum_{i=1}^c \sum_{\mathbf{x} \in \chi_i} (\mathbf{x} - \mathbf{m}_i)^T (\mathbf{x} - \mathbf{m}_i)$$

n : 全データ数

c : クラス数

χ_i : クラス i のデータ

\mathbf{m}_i : クラス i の平均

- クラス間分散

$$\sigma_B^2 = \frac{1}{n} \sum_{i=1}^c n_i (\mathbf{m}_i - \mathbf{m})^T (\mathbf{m}_i - \mathbf{m})$$

\mathbf{m} : 全データの平均

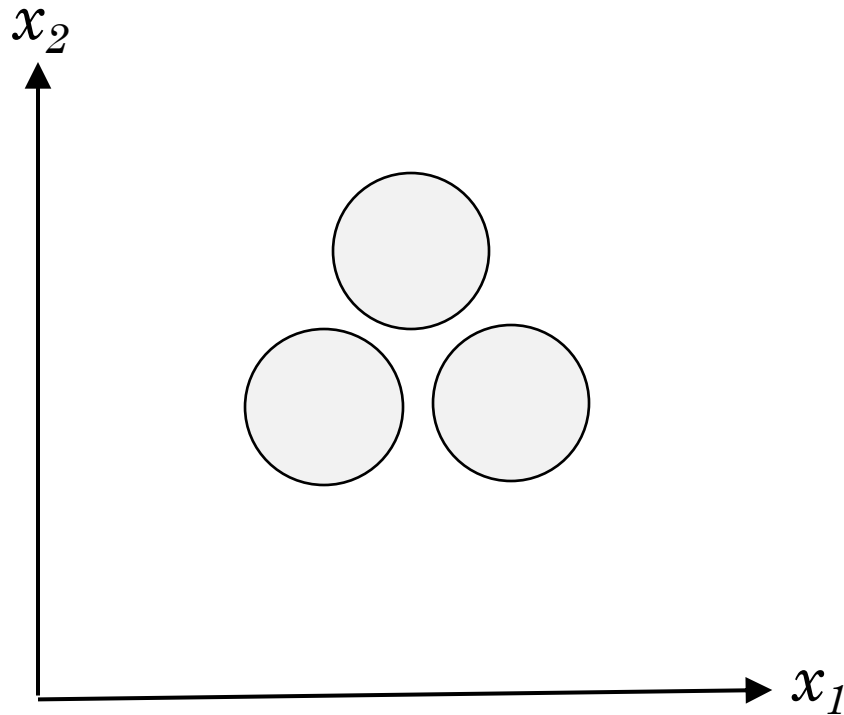
n_i : クラス i のデータ数

- クラス内分散・クラス間分散比 (大きいほど良い)

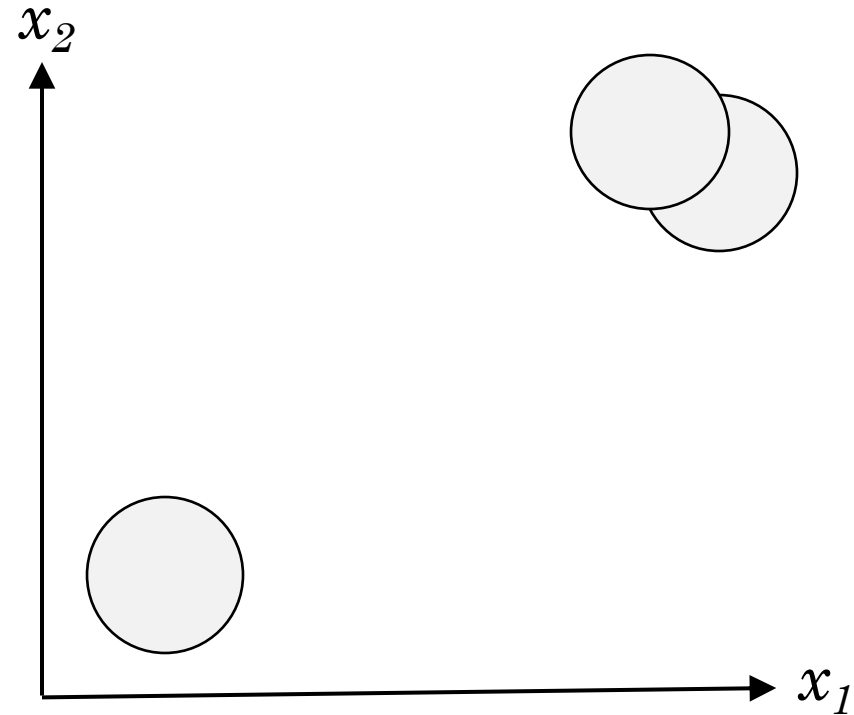
$$J_\sigma = \frac{\sigma_B^2}{\sigma_W^2}$$

定量評価

- 多クラスのクラス内分散・クラス間分散比
 - 分布の重なりを考慮できないので、あまりよい評価尺度とはいえない



(a) クラス間分散：小



(b) クラス間分散：大

第2週の実験

- 特徴の評価と選択

クラス内分散・クラス間分散比を求めるGoogle Colaboratoryのコードを作成し、識別に有効であると思われる2次元特徴の組み合わせを3つ程度求めよ（定量評価）

- 上で求めた組み合わせについて、2次元散布図を出力するGoogle Colaboratoryのコードを作成し、識別に最も有効だと考えられる特徴の組み合わせを求めよ（定性評価）
- （発展課題） 特徴ベクトルの次元数を3次元とし、評価の高い組み合わせを求めよ