

# プロジェクト実習 III

## パターン認識

### - 第4週目 -

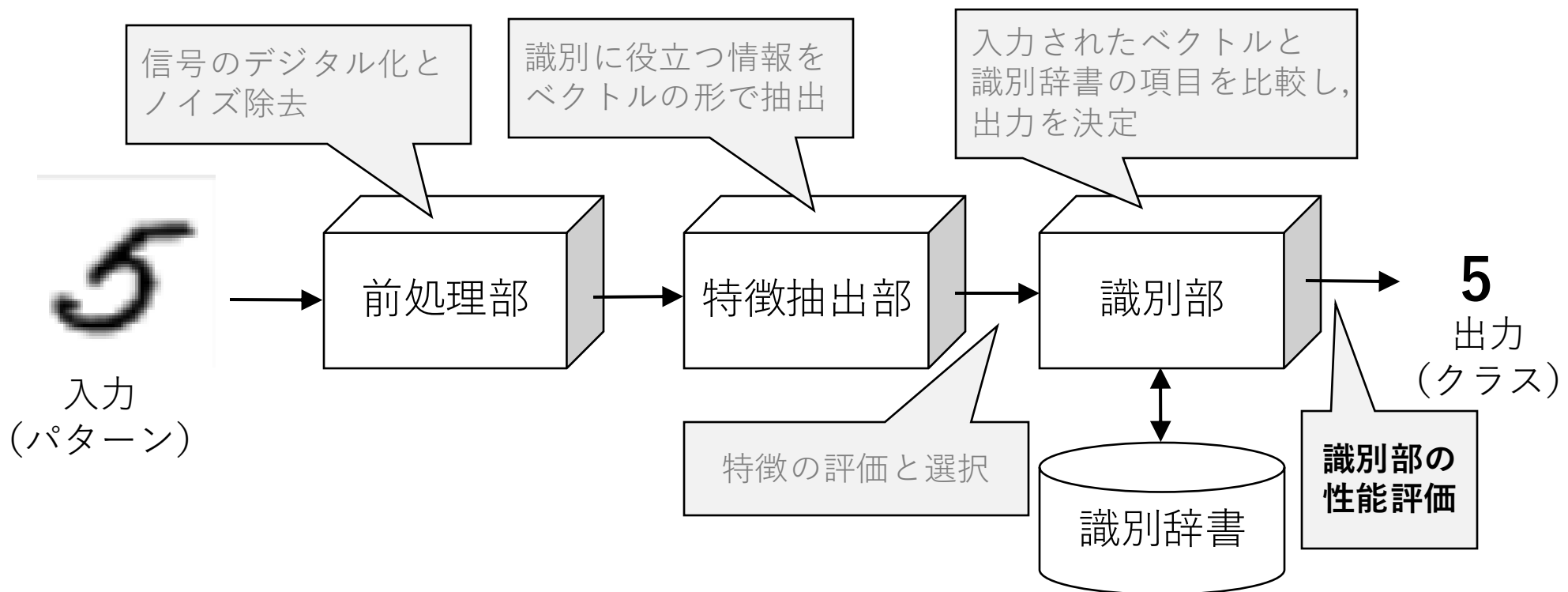
担当：崔恩瀨

# 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4週の実験



# 識別部の実装（第3週の実験結果）

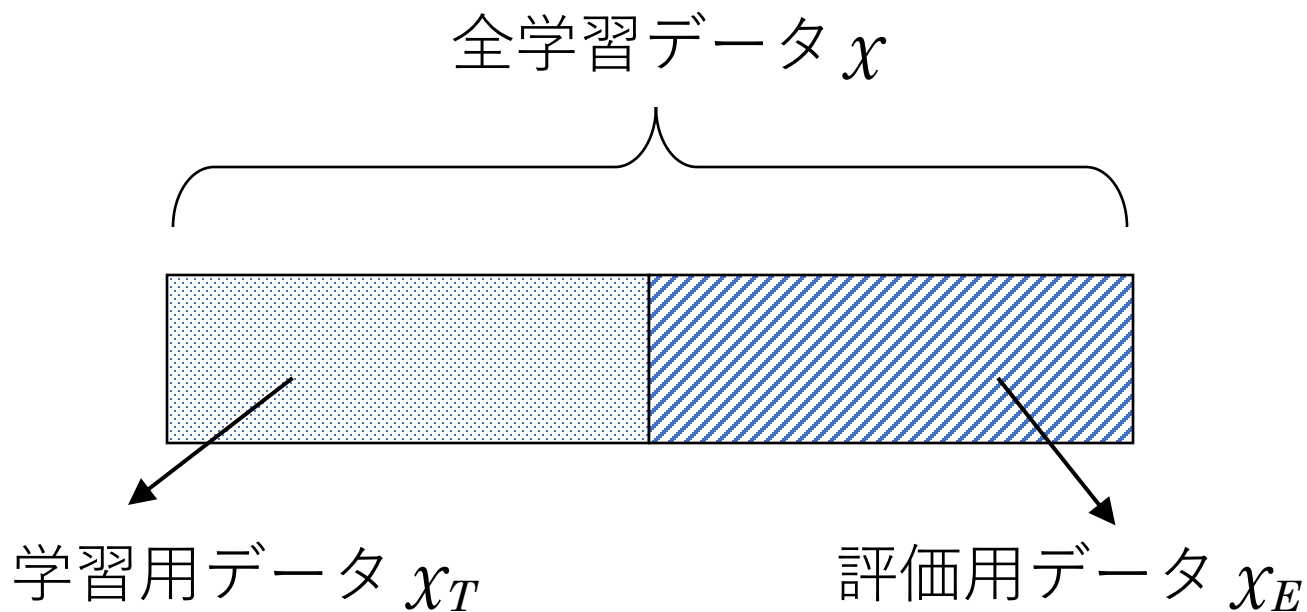
- 識別部が正しく実装されていることの確認
  - 高い性能を示した
    - おそらく識別率は95%前後となったはず
- この性能を信用して良いのか
  - 実運用では学習データとは異なるデータが入力される
  - 学習データを使って評価してよいなら， 識別率100%の識別器が容易に作れる

# 識別部の評価（第4週の実験内容）

- 実装した識別部の評価
  - 未知データに対する性能を見積もりたい
- 評価法
  - 分割学習法
  - 交差確認法

# 分割学習法

- 手順
  - 全学習データ  $\mathcal{X}$  を学習用データ集合  $\mathcal{X}_T$  と評価用データ集合  $\mathcal{X}_E$  に分割する
  - $\mathcal{X}_T$  を用いて識別機を設計し,  $\mathcal{X}_E$  を用いて識別率を推定する



# 分割学習法

- 利点
  - 評価が容易
- 欠点
  - 学習に用いるデータ数が減るので、識別性能が劣化する
  - 評価に用いるデータ数が少ない場合、識別率の推定精度は良くない

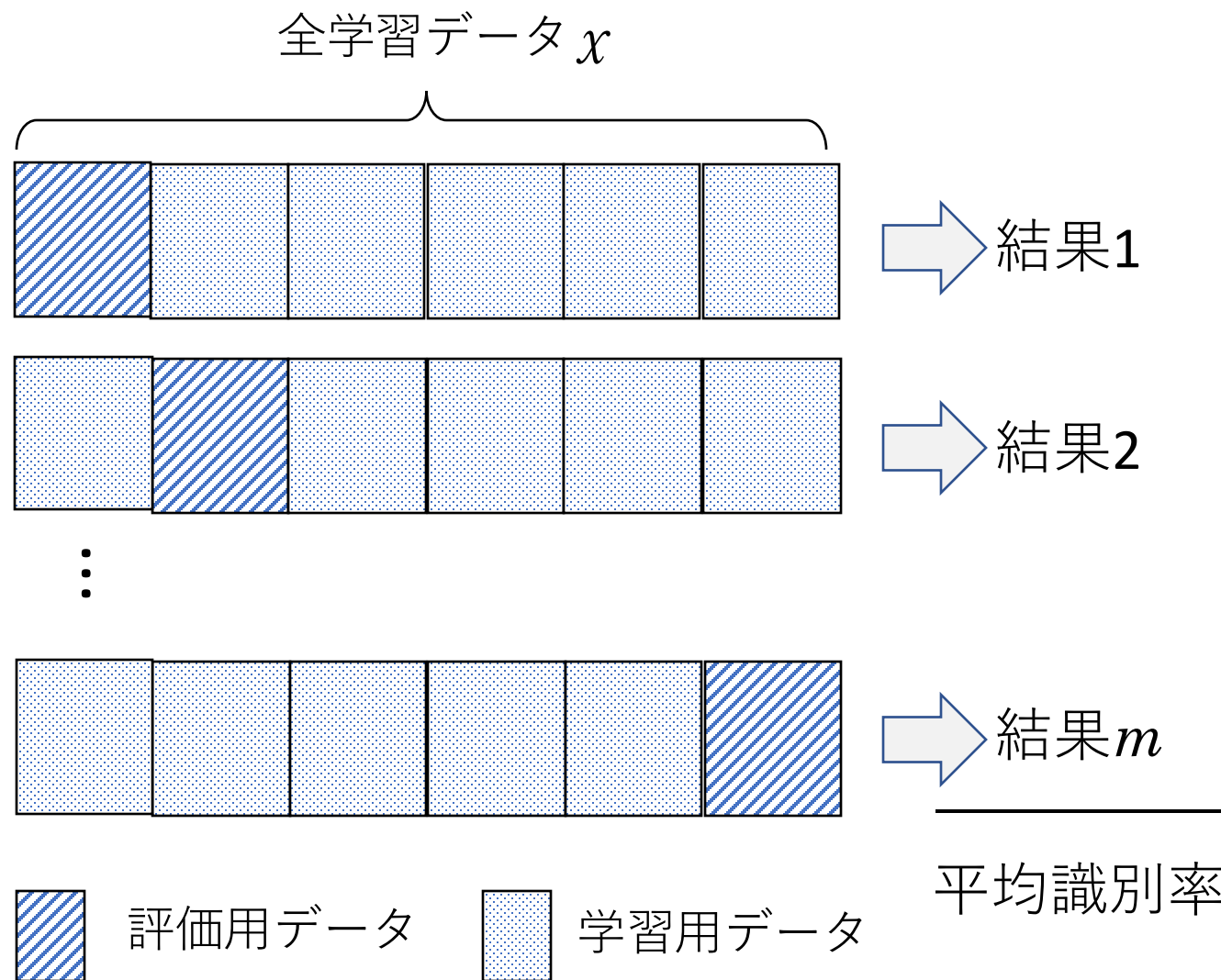
# 交差確認法

- 手順

1.  $x$  を  $m$  個のグループ  $x_1, \dots, x_m$  に分割する
2.  $x_i$  を除いた  $(m-1)$  個のグループで学習し,  $x_i$  を用いて識別率を算出する
3. この手順をすべての  $i$  について行い,  $m$  個の識別率の平均を識別率の推定値とする



# 交差確認法



# 交差確認法

- 利点
  - 分割学習法に比べ，識別率の推定精度は高い
- 欠点
  - 評価に時間がかかる
  - 分割数が少ない場合，分割法が異なると評価値が大きくぶれる

# 第4週の実験

- 識別部の評価
  - 第3週に作成した識別器について分割学習法，交差確認法(10分割)のそれぞれで評価を行うコードをGoogle Colaboratoryで作成せよ
  - 分割学習法，交差確認法それぞれの評価結果を比較し，得失を論ぜよ
  - （発展課題）最近傍決定則の発展として，近傍 $k$ 個のデータが属するクラスの多数決を識別結果とする $k$ -NN法を実装し，異なる $k$ について交差確認法で性能を評価せよ
  - （発展課題）交差確認法においてデータ分割の方法を乱数を用いて変化させ，複数回実行して平均値を求める方法を実装せよ