

5章のストーリー

- 清原は、ディープラーニングを使って糖尿病診断システムに再チャレンジしようとする
- さやかは識別精度を向上させる別の方法としてアンサンブル学習を教える

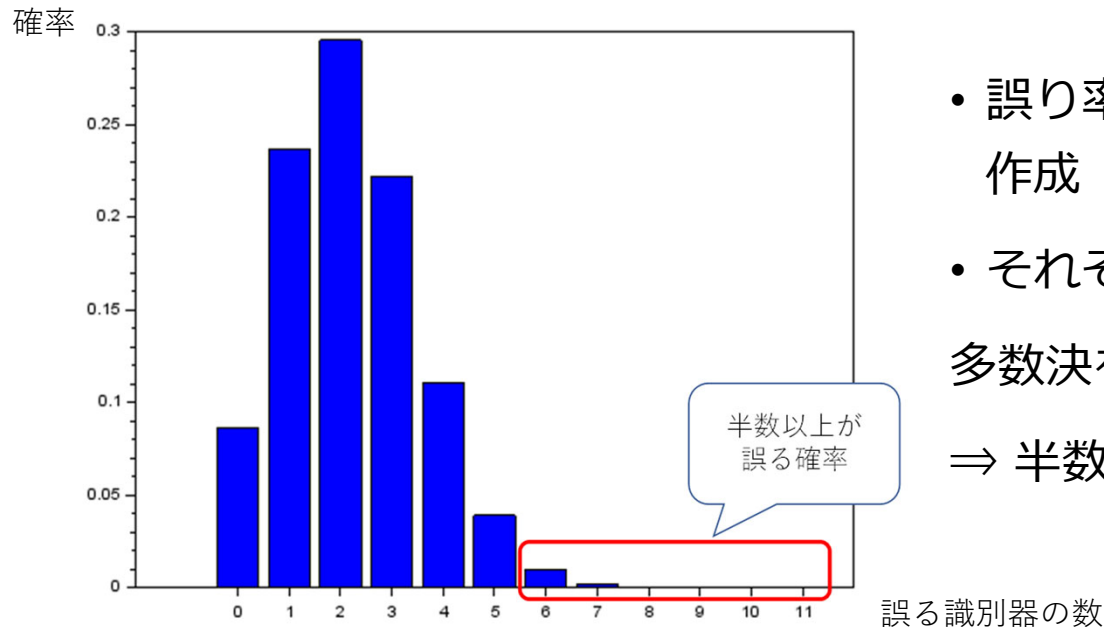
アンサンブル学習 (5章)



p.146 2コマ目

アンサンブル学習

- アンサンブル学習とは
 - 識別器を複数組み合わせ、それらの結果を統合することで、個々の識別器よりも性能を向上させる方法
- なぜ複数の識別器で性能が向上するのか



- 誤り率20%の識別器を11個作成
 - それぞれが誤るのは独立
- 多数決を取った場合
- ⇒ 半数以上が誤るのは1.2%

アンサンブル学習

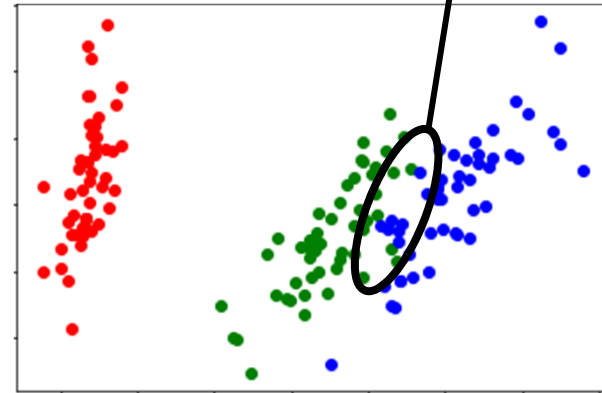
- ここまでの議論の非現実的なところ

「それぞれの識別器の誤りが独立」

⇒ データの誤りやすさに差はない ×

識別面付近のデータなど、
普通は成立しない

多くの識別器が誤る



- アンサンブル学習の目標

- なるべく異なる振る舞いをする識別器を作成する

アンサンブル学習

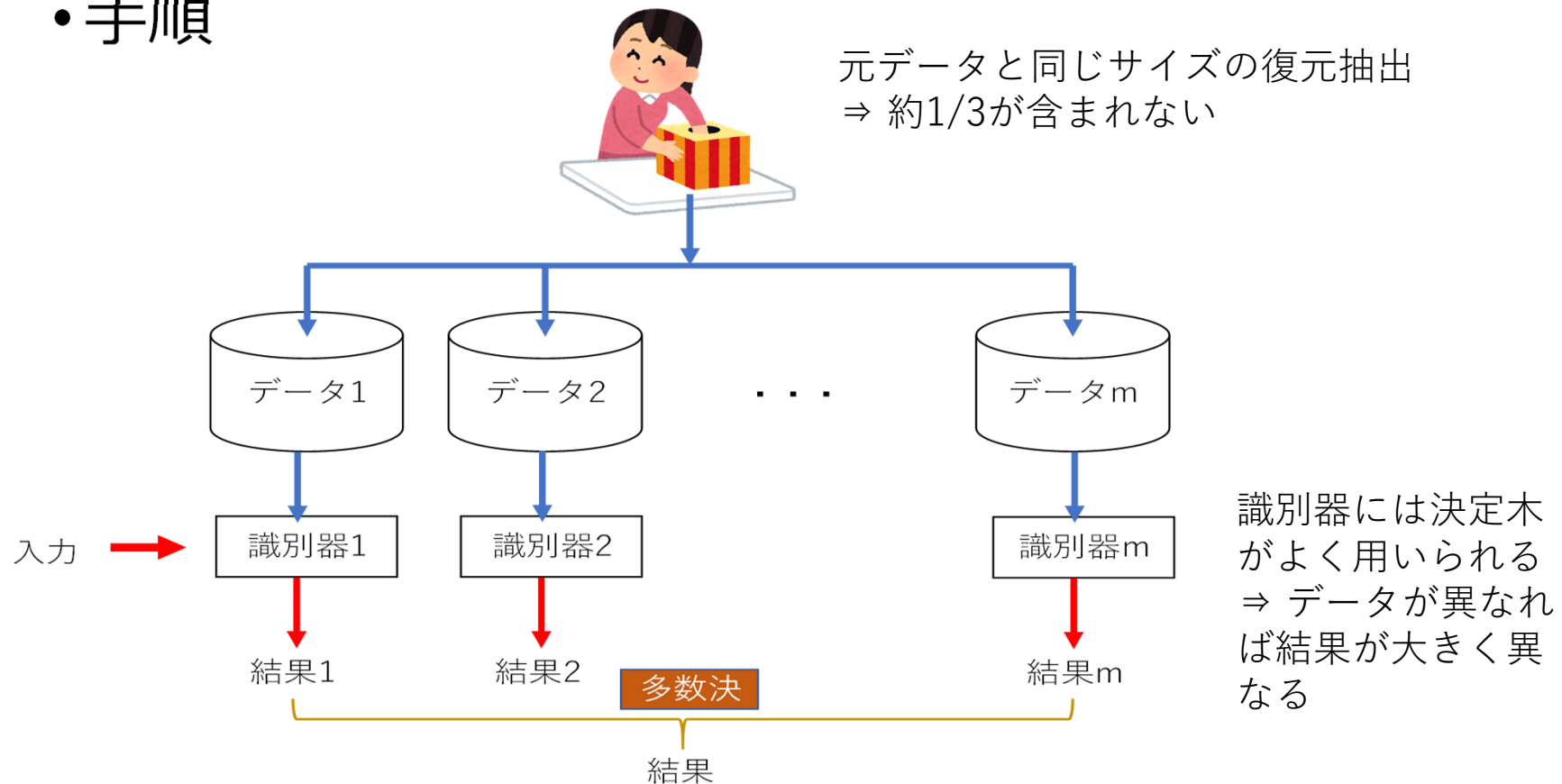
- アンサンブル学習の手法
 - バギング
 - ランダムフォレスト
 - ブースティング

バギング

- バギングのアイデア

- 異なる学習データから作成された識別器は異なる

- 手順

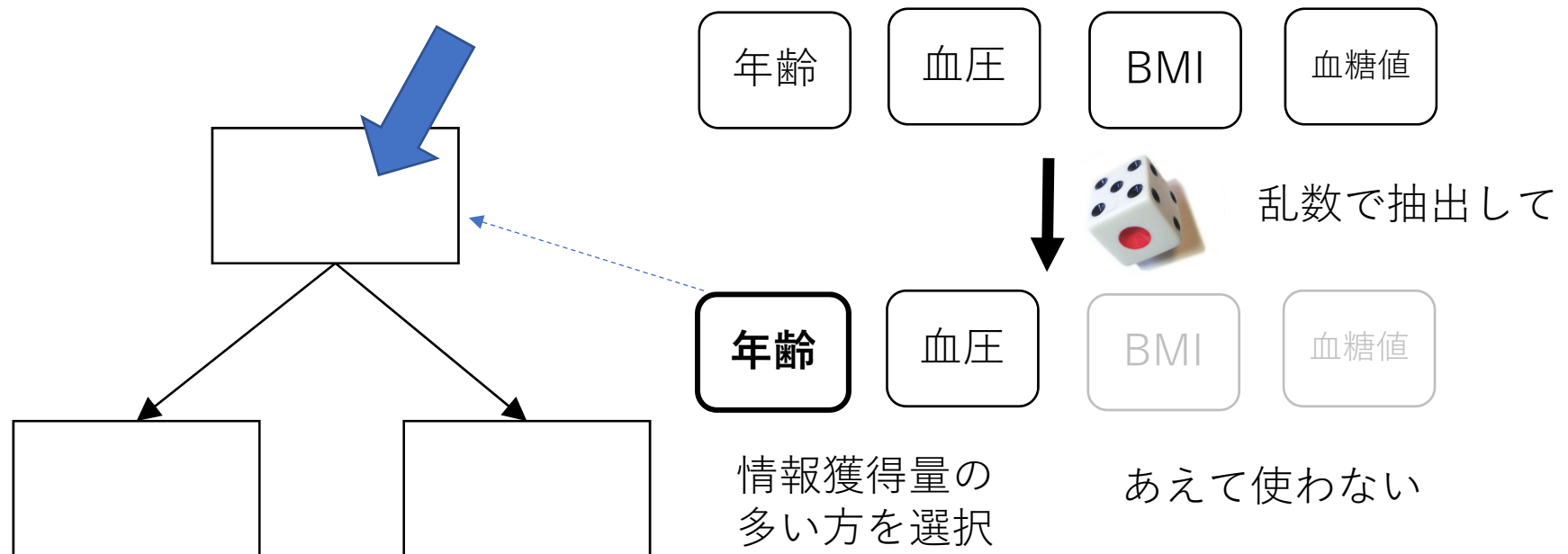


ランダムフォレスト

- ランダムフォレストのアイディア
 - バギング+識別器を作成する毎に異なる特徴を用いる
ことで異なった識別器を複数作成する

- 手順

この分岐条件を選ぶときに...

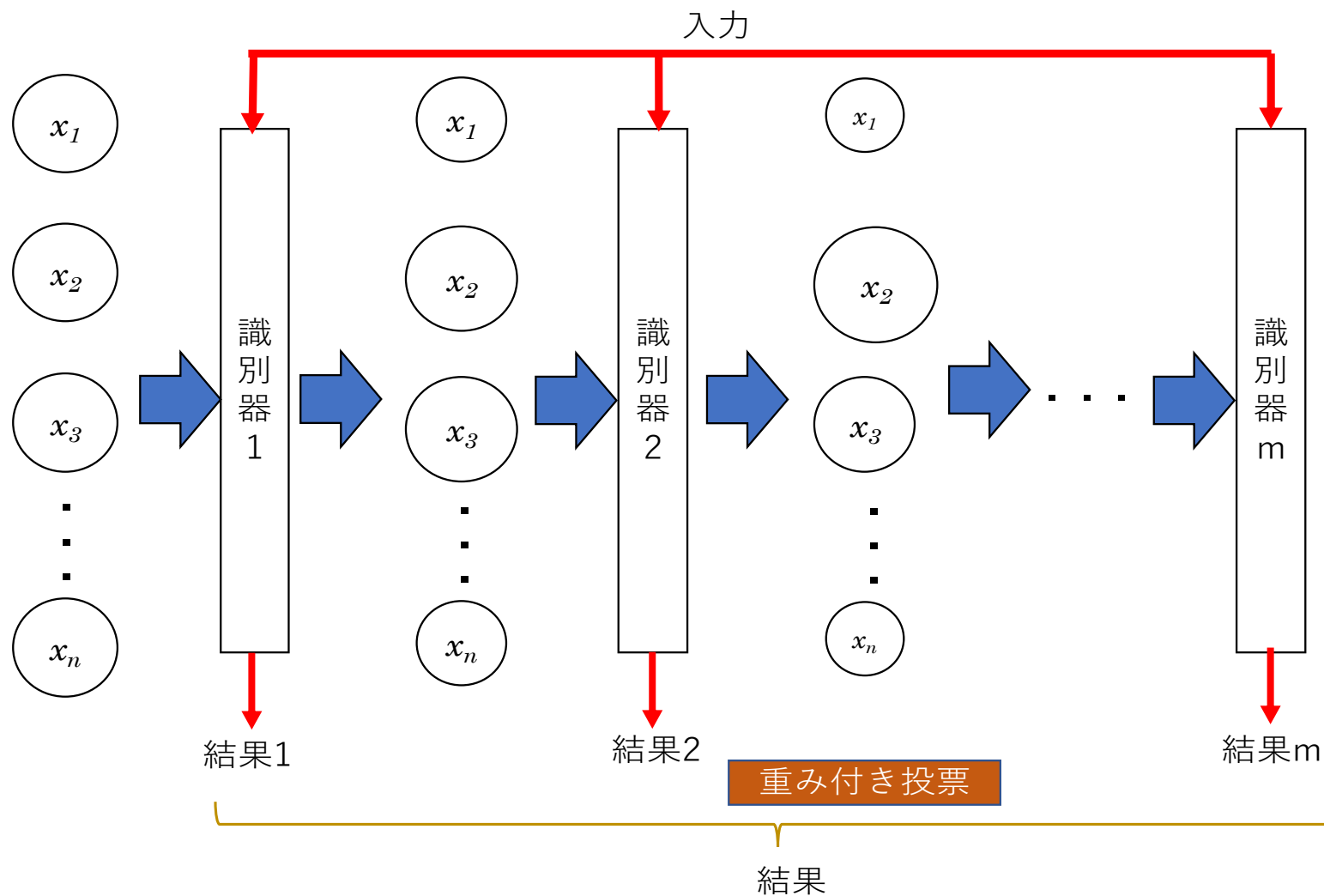


ブースティング

- ブースティングのアイデア
 - 現在の識別器が誤識別を起こすデータを正しく識別する識別器を逐次的に追加
 - 過学習とならないように、識別器として浅い決定木を用いることが多い

AdaBoost

- 前段の識別器が誤ったデータの重みを重くする



勾配ブースティング

- 損失が最小となるような識別器を逐次加える

