



# AI WorkShop

金融データを扱った機械学習の演習  
フィナンシャル機械学習 第6章 アンサンブル法

2021/12/17

土田晃司

# 金融データを扱った機械学習の演習

## フィナンシャル機械学習 第6章 アンサンブル法

### 目次

1. 機械学習モデルの3つの誤り
2. アンサンブル法の手法 1
3. アンサンブル法の手法 2
4. バギングとブースティングの比較
5. まとめ
6. 演習問題

# 1. 機械学習モデルの3つの誤り

## 1. バイアス:過度

非現実的な仮定をすることによる誤りで、アンダーフィットしていることになる。

## 2. バリアンス:感度

訓練データセットにおけるわずかな変化に対する感度が高いことによる誤りで、オーバーフィットしていることになる。

## 3. ノイズ:誤差

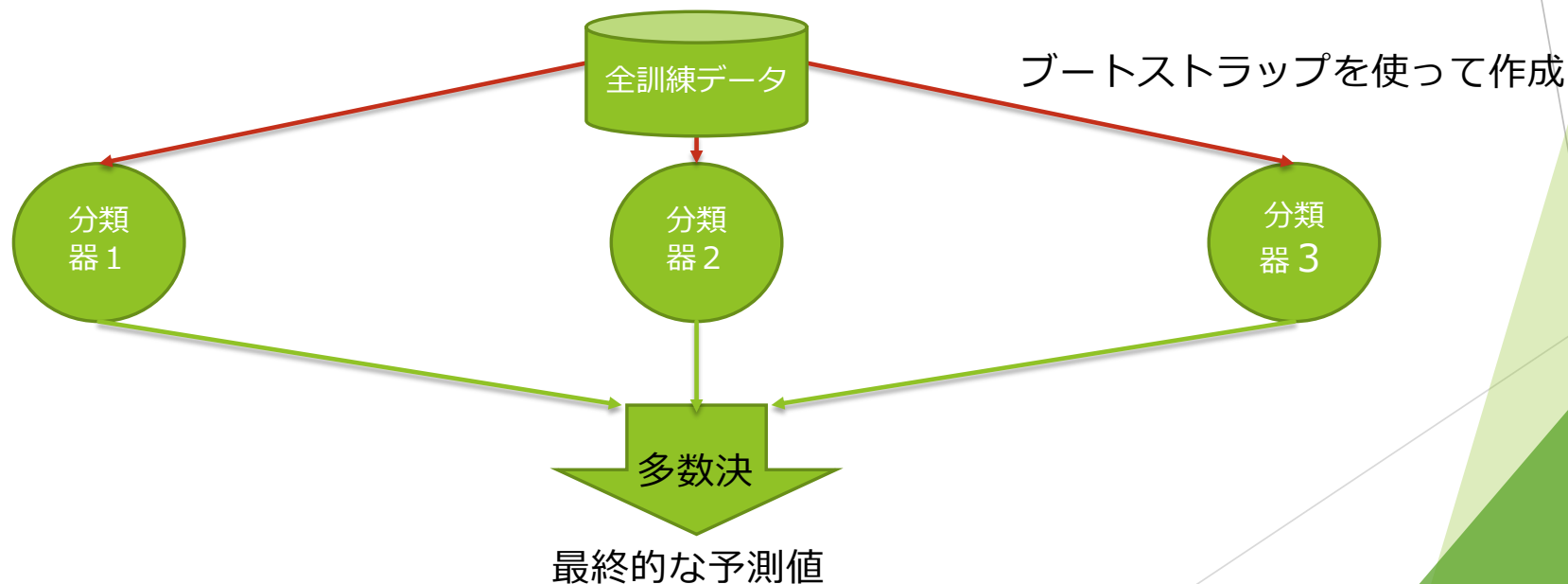
予測不可能な変化や測定誤差などの観測値のバリアンスによる誤りで、既約誤差になる。

## 2. アンサンブル法の手法 1

### ▶ バギング(ブートストラップアグリゲーション)

弱分類器を独立に多数作り、それらの結果による多数決を行なって予測するモデル。各弱分類器の正解率を上げることはできない。(ランダムフォレスト)

バリエアンスを減らす効果があり、オーバーフィットへの対処ができる。冗長なデータに対しては、過大な正解率を示す場合がある

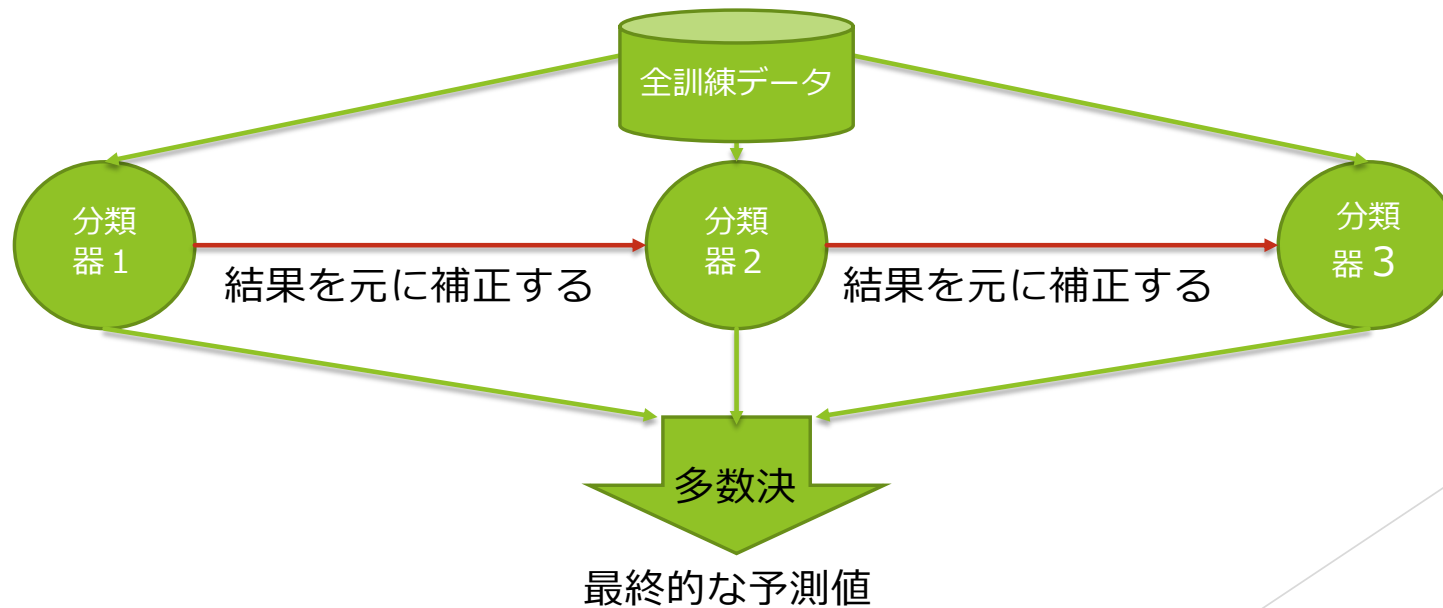


## 3. アンサンブル法の手法 2

### ▶ ブースティング

順番に弱分類器を構成し、それらの結果による多数決を行なって予測するモデル。各弱分類器の正解率を上げることができる。（AdaBoost、XGBoost）

バイアスとバリエーション両方を減らすことができるが、バイアスを減らすとオーバーフィットの可能性が増す



## 4. バギングとブースティングの比較

- ▶ バギングは並列処理、ブースティングは逐次実行。
- ▶ バギングは単一分類器の結果を出力、ブースティングは各分類器の加重平均になる。
- ▶ バギングはバリエーションのみ削減できる、ブースティングはバイアスとバリエーション両方を削減できる。

## 5. まとめ

- ▶ 金融データはノイズが少ないので、バイアスとバリエーションに注意すればよい
- ▶ 金融データを扱う場合は、オーバフィットすることが好ましくないなので、バギングを利用する方がよい
- ▶ バギングの方が並列で処理できるので、学習が早く完了する。
- ▶ 演習ではバギングとブースティングを比べます。
- ▶ Deeplarningの学習の流れをおさらいします。