

## 第7章

- 1 もっと良い結果が得られると考えられる。独立している場合、多数決方式をとるとすると、3つ以上が誤った結果を返す確率は $(0.05)^3 + (0.05)^4 + (0.05)^5 = 0.0013128125$ なので約99%の適合率を達成できると考えられる。モデル同士の独立度合が低い場合はより良い結果を返さないこともある。
- 2 それぞれの分類器の分類結果だけを見るかその確率までを考慮するかの違い
- 3 バギングアンサンブル、ブースティングアンサンブル、ランダムフォレストはそれぞれの分類器を別のサーバーで訓練することができるのでスピードを上げることができる。ブースティングアンサンブルは分類器の訓練に前の分類器の情報が必要なため、複数のサーバーでスピードを上げることは難しい。スタッキングアンサンブルでは同じ層の分類器を訓練するのは複数のサーバーを用いることが可能だが、ある層の訓練にはその前の層の訓練が終わっている必要があるため、完全な並列化はできない。
- 4 新たにテストデータを用意しなくてよいので訓練に用いることのできるインスタンス数が増えること。
- 5 通常のランダムフォレストより訓練は速いが、推定の時間は変わらない。わざと無作為にすることでそれぞれの分類器の独立性を確保して、分散を下げています。
- 6 分類器の数を増やすか、元の分類器の正則化ハイパーパラメータを下げる。
- 7 学習率は下げるべき。