7章　アンサンブル学習とランダムフォレスト

1. もっと良い結果が得られる可能性はある。5つの異なるモデルから得られた結果の多数決をとり、選ばれた結果を全体の出力とするようにアンサンブルを作ればよい。
2. ハード投票分類器は、アンサンブルを構成する学習器それぞれの出力の多数決をとって、その結果をアンサンブルの出力とするような方式である。一方で、ソフト投票分類器は、学習器それぞれが出力する、インスタンスがクラスに属する確率の平均値をとり、それが最も大きいクラスを出力とするような方式である。
3. バギングは並列処理が可能であるので、スピードが上がる。ペースティングアンサンブル、ランダムフォレスト、スタッキングアンサンブルでもスピードが上がる一方、ブースティングアンサンブルは逐次的な手法であるため、スピードは上がらない。
4. 検証セットを別に用意せずとも、訓練セットのうち訓練に使われていないインスタンスを用いて検証できる点。
5. ランダムフォレストはノードの分割の際、特徴量の無作為なサブセットから最良の特徴量を探して適切なしきい値を設定する一方で、Extra-Treeではしきい値を無作為に設定するという違いに起因する。このことには、バイアスを少し上げて分散を下げることになり、過学習のリスクを下げるという意味がある。また、Extra-Treeは、特徴量の最良のしきい値を探すという時間のかかるタスクを行う必要がないため、ランダムフォレストよりも訓練が速くなる。
6. n_estimatorsの値をふやす。
7. 学習率を下げるべきである。
8. ファイル「code_7.ipynb」参照のこと。まず、ハード投票分類器を調べると、正解率がExtra-Tree 0.9748、ランダムフォレスト分類器 0.9709、SVM分類器 0.9802に対して0.9779であり、SVM分類器に性能が及ばなかった。次に、ソフト投票分類器は正解率が0.981で、これは単体の分類器よりも若干高い。テストセットでは、ソフト投票分類器が0.9784であった。
9. ファイル「code_7.ipynb」参照のこと。ブレンダを用いた場合、正解率は0.9756であった。これは、前問の正解率と比べて有意な差はない。