

6章 決定木

1. $\log_2(10^6) \doteq 20$ 回の分岐でインスタンスごとに葉ノードができるので、深さは20。
2. 常に低い
3. ノードの過度な分割を防ぐことができるため、よい
4. 入力特徴量は関係がない。
5. 決定木の訓練の計算量は $O(n \times m \times \log_2 m)$ である。よって、訓練セットのインスタンス数が10倍になったとき、訓練時間はおよそ $10 \times \log_2(10^7) / \log_2(10^6) = 35/3$ 倍になる。したがって、求める時間は、1時間 $\times 35/3 = 11$ 時間40分
6. 上がらない。プレソートによって訓練が高速化するのは、インスタンス数が数千個未満の場合である。
7. ファイル「code_6_7.ipynb」参照のこと。グリッドサーチで、max_leaf_nodesとmax_depthを調べたところ、max_leaf_nodes=4、max_depth=2が最適であると分かった。この結果、検証セットでは86.9%の正解率が得られた。
8. ファイル「code_6_8.ipynb」参照のこと。テキストの指示通りに森を育てたつもりであるが、正解率は86.7%で、単体の決定木の86.9%を上回れなかった。