

Classification（分類）

モデル	Pros	Cons
ロジスティック回帰	確率的な考え方であると同時に、変数にどれほどのインパクトがあるかを評価することができる	ロジスティック回帰の前提を満たす必要がある
K近傍法	直観的に理解でき、かつ実行が早い	Kの数をこちらで決める必要がある
サポートベクトルマシン	外れ値に影響を受けず、過学習もしづらい	非線形の場合や、特徴量が多い場合には適さない
カーネルSVM	非線形の問題に対して高いパフォーマンスを発揮し、かつ外れ値に影響を受けず、過学習もしづらい	特徴量が多い場合は最適な選択肢ではなくなる
ナイーブベイズ	外れ値に影響を受けず、非線形の問題にも対処可能。確率的な考え方をする	変数が独立していることが前提となる
分類木	フィーチャースケーリングが不要なく、線形/非線形のどちらにも適用可能	データが少ないと良い結果が得られない。過学習がおこりがち
ランダムフォレスト	正確性が高く、非線形を含めて多くの問題に対処可能	直感的な解釈が難しく、かつ過学習しやすい。木の数を決める必要がある