

目次

第 1 章	予測に欠かせない二つのアルゴリズム	1
1.1	なぜこの二つのアルゴリズムは便利なのか？	1
1.2	罰則付き線形回帰とは？	6
1.3	アンサンブル法とは？	8
1.4	使用するアルゴリズムの決め方	10
1.5	予測モデルを構築するプロセス	12
1.5.1	機械学習問題の構成	14
1.5.2	特徴抽出と特徴エンジニアリング	15
1.5.3	学習したモデルの精度を決める	16
1.6	章の内容と依存関係	17
1.7	本章のまとめ	19
	参考文献	19
第 2 章	データを理解することによって問題を理解する	21
2.1	新しい問題の構造	21
2.1.1	属性とラベルの型の違いがモデル選択を促進させる	23
2.1.2	新しいデータセットに対する気づき	25
2.2	分類問題——音波探知機を使って不発の機雷を見つける	26
2.2.1	岩と機雷のデータセットの物理的特徴	26
2.2.2	岩と機雷のデータセットの要約統計量	29
2.2.3	QQ プロットを使って外れ値を可視化する	32
2.2.4	質的変数の統計的特徴	34
2.2.5	Pandas パッケージを使って、岩と機雷のデータセットを要約する	35
2.3	岩と機雷のデータセットの性質を可視化する	37
2.3.1	平行座標プロットで可視化する	37
2.3.2	属性とラベルの関係を可視化する	39
2.3.3	ヒートマップを使って属性とラベルの相関関係を可視化する	45

2.3.4	岩と機雷のデータセットを理解するためのデータ要約プロセス	47
2.4	質的データで量的な予測を行う——アワビは何才か?	47
2.4.1	回帰問題に対する平行座標プロット——アワビの問題の変数関係を可視化する	52
2.4.2	回帰問題に対する相関係数のヒートマップ——アワビの問題の相関関係を可視化する	56
2.5	実数を持つ属性を使った実数の予測——ワインの味をどのように計算するか	58
2.6	多クラス分類問題——どんなタイプのガラスか?	64
2.7	本章のまとめ	69
	参考文献	69

第3章 予測モデルの構築——精度、複雑さ、データ量のバランス 71

3.1	基本的な問題——関数近似を理解する	71
3.1.1	学習用データを使用する	72
3.1.2	予測モデルの精度を測る	74
3.2	アルゴリズムの選択を左右する要素と精度——複雑さとデータ	75
3.2.1	単純な問題と複雑な問題の対比	75
3.2.2	単純なモデルと複雑なモデルの対比	77
3.2.3	予測アルゴリズムの精度を左右する要素	80
3.2.4	アルゴリズムの選択——線形か非線形か	82
3.3	予測モデルの精度を測る	82
3.3.1	さまざまなタイプの問題の精度を測る	83
3.3.2	デプロイしたモデルの精度を評価する	93
3.4	モデルとデータの調和を生み出す	95
3.4.1	問題の複雑さ、モデルの複雑さ、データセットのサイズのバランスで、モデルを選択する	96
3.4.2	過学習をコントロールするために前進ステップワイズ回帰を使う	97
3.4.3	予測モデルの評価と理解	102
3.4.4	罰則付き回帰係数で過学習をコントロール——リッジ回帰	104
3.5	本章のまとめ	112
	参考文献	113

第 4 章	罰則付き線形回帰	115
4.1	罰則付き線形回帰が有用である理由	116
4.1.1	学習の高速性	116
4.1.2	変数の重要度	116
4.1.3	実行の高速性（デプロイ後）	117
4.1.4	信頼性の高い精度	117
4.1.5	疎な解	117
4.1.6	線形モデルを要求する問題	118
4.1.7	アンサンブル手法を使う場合	118
4.2	罰則付き線形回帰——線形回帰問題に罰則項を加えて精度を高める	118
4.2.1	線形モデルの学習——誤差の最小化とその周辺	120
4.3	罰則付き線形回帰問題の解法	125
4.3.1	LARS, および LARS と前進ステップワイズ回帰の関係	125
4.3.2	一般的で高速な手法“glmnet”の利用	138
4.4	数値入力を伴う線形回帰への拡張	145
4.4.1	罰則付き回帰による分類問題の解き方	145
4.4.2	二つ以上の目的変数を有する分類問題	149
4.4.3	基底展開の理解——非線形問題に対する線形手法の適用	150
4.4.4	非数値説明変数の線形モデルへの統合	152
4.5	本章のまとめ	157
	参考文献	158
第 5 章	罰則付き線形回帰を用いた予測モデル構築	159
5.1	罰則付き線形回帰の Python パッケージ	160
5.2	多変数回帰——ワインの味の予測	161
5.2.1	ワインの味の予測モデルの構築と精度評価	162
5.2.2	デプロイ前のデータセット全体に対する学習	167
5.3	2 クラス分類——罰則付き線形回帰を用いた不発弾の検出	176
5.3.1	デプロイ用の岩と機雷の分類器	186
5.4	多クラス分類——犯罪現場のガラスサンプルの分類	199
5.5	本章のまとめ	204
	参考文献	205

第 6 章	アンサンブル学習	207
6.1	決定木	208
6.1.1	決定木はどのように予測を行うか	210
6.1.2	決定木をどのように構築するか	211
6.1.3	決定木の学習は分割点の設定と等価である	213
6.1.4	決定木の過学習	217
6.1.5	分類問題とカテゴリカルデータへの拡張	221
6.2	ブートストラップ集約——バギング	222
6.2.1	バギングアルゴリズムはどのように動くか	222
6.2.2	バギングのまとめ	232
6.3	勾配ブースティング	232
6.3.1	勾配ブースティングアルゴリズムの基本原理	233
6.3.2	勾配ブースティングから最も良い精度を得る方法	236
6.3.3	多変量データへの勾配ブースティングの適用	239
6.3.4	勾配ブースティングのまとめ	242
6.4	ランダムフォレスト	243
6.4.1	ランダムフォレスト——バギングとランダムな変数を持つデータ 小集団	246
6.4.2	ランダムフォレストの精度について	247
6.4.3	ランダムフォレストのまとめ	248
6.5	本章のまとめ	249
	参考文献	249
第 7 章	アンサンブル学習のモデル構築	251
7.1	Python の Ensemble パッケージによる回帰問題の解決	251
7.1.1	ランダムフォレストモデルによるワインの品質予測モデルの構築	252
7.1.2	ワインの品質データセットへの勾配ブースティングの適用	259
7.2	ワインの品質データセットへのバギングの適用	266
7.3	非数値変数を含むデータに対するアンサンブル学習モデルの構築	271
7.3.1	ランダムフォレスト回帰に適用するためのアワビの性別データの 入力法	271
7.3.2	非数値変数を含むデータにおける精度と重要な変数の評価	274
7.3.3	勾配ブースティングに適用するためのアワビの性別データの 入力法	275

7.3.4	非数値変数を含むデータに勾配ブースティングを適用したときの 精度と重要な変数の評価	278
7.4	2クラス分類問題へのアンサンブル学習の適用	280
7.4.1	アンサンブル学習を用いた機雷の識別	281
7.4.2	機雷の識別のためのランダムフォレストモデルの構築	282
7.4.3	ランダムフォレスト学習器による分類精度の検証	286
7.4.4	機雷の識別のための勾配ブースティングの構築	288
7.4.5	勾配ブースティング学習器による分類精度の検証	294
7.5	多クラス分類問題へのアンサンブル学習の適用	297
7.5.1	ランダムフォレストによるガラスの分類	298
7.5.2	クラス不均衡への対応	301
7.5.3	勾配ブースティングによるガラスの分類	302
7.5.4	勾配ブースティングの基本学習器にランダムフォレストを使用す る場合の利点	307
7.6	アルゴリズムの比較	309
7.7	本章のまとめ	310
	参考文献	311

索引 313

著者と技術編集者について 316