Table of Contents

Pref	ace Gradient Boosting	Model selection	1
Cha	pter 1: A Process for Success	Summary	9
68	The process	pter 4: Advanced Feature Sel	10
	Business understanding		11
	Identifying the business objective		12
	Assessing the situation		13
	Determining the analytical goals		13
	Producing a project plan		13
	Data understanding		14
	Data preparation		14
	Modeling		15
	Evaluation		16
	Deployment		16
	Algorithm flowchart		17
	Summary		22
Cha	pter 2: Linear Regression - The Blockin	ng and Tackling of Machine	
	rning example of deep learning	Model selection	23
711	Univariate linear regression		24
	Business understanding		27
	Multivariate linear regression		33
	Business understanding		33
	Data understanding and preparation		34
	Modeling and evaluation		37
	Other linear model considerations		51
	Qualitative features		51
	Interaction terms		53
	Summary		55
Cha	pter 3: Logistic Regression and Discri	minant Analysis	57
139	Classification methods and linear regre		58
	Logistic regression		58
	Business understanding		59
	Data understanding and preparation		60
	Modeling and evaluation		65

	The logistic regression model		66
	Logistic regression with cross-validation		69
	Discriminant analysis overview		72
	Discriminant analysis application		75
	Multivariate Adaptive Regression Splin	nes (MARS)	78
	Model selection		84
	Summary		8
Cha	pter 4: Advanced Feature Selection in	Linear Models	89
	Regularization in a nutshell		90
	Ridge regression		9
	LASSO		9
	Elastic net		92
	Business case		92
	Business understanding		92
	Data understanding and preparation		9:
	Modeling and evaluation		99
	Best subsets		99
	Ridge regression		103
	LASSO		108
	Elastic net		11
	Cross-validation with glmnet		114
	Model selection		116
	Regularization and classification		11
	Logistic regression example		11
	Summary		120
Cha	pter 5: More Classification Technique	s - K-Nearest Neighbors and	
	nort Vector Machines	Data hac polipactachau atc	12
37	K-nearest neighbors	Modeling and evaluation	122
			123
	Business case		12
	Business understanding		12
	Data understanding and preparation		128
	Madaling and avaluation		
	KNN modeling		134
	SVM modeling		139
	Model selection	Logistic regression	14:
	Feature selection for SVMs		14
	Summary		147
Cha	pter 6: Classification and Regression	Trees taulave bna gnileboM	149

Understanding the regression trees Classification trees Random forest Random forest Gradient boosting Business case Modeling and evaluation Regression tree Classification tree Random forest regression Random forest regression Random forest classification Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary Chapter 7: Neural Networks and Deep Learning Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning Data upload to H2O Create train and test datasets
Random forest Gradient boosting Business case Modeling and evaluation Regression tree Classification tree Random forest classification Random forest classification Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary Chapter 7: Neural Networks and Deep Learning Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning Pata upload to H2O Create train and test datasets
Gradient boosting Business case Modeling and evaluation Regression tree 154 Classification tree 158 Random forest regression Random forest classification Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary Chapter 7: Neural Networks and Deep Learning 177 Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning Data upload to H2O Create train and test datasets 154 155 160 160 160 160 161 160 160 160 160 160
Business case Modeling and evaluation Regression tree Classification tree Random forest regression Random forest classification Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary Chapter 7: Neural Networks and Deep Learning Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning Data upload to H2O Create train and test datasets 154 Modeling and evaluation 155 Modeling and evaluation 156 157 158 Random forest regression 160 163 163 164 165 177 177 177 177 177 178 177 177
Modeling and evaluation 154 Regression tree 154 Classification tree 158 Random forest regression 160 Random forest classification 163 Extreme gradient boosting - classification 167 Model selection 172 Feature Selection with random forests 172 Summary 175 Chapter 7: Neural Networks and Deep Learning 177 Introduction to neural networks 177 Deep learning, a not-so-deep overview 182 Deep learning resources and advanced methods 184 Business understanding 186 Data understanding and preparation 187 Modeling and evaluation 192 An example of deep learning 197 H2O background 198 Data upload to H2O 198 Create train and test datasets 200
Modeling and evaluation Regression tree Classification tree Classification tree Random forest regression Random forest classification Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary 172 Summary 175 Chapter 7: Neural Networks and Deep Learning 177 Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning Data upload to H2O Create train and test datasets 200
Classification tree
Random forest regression Random forest regression Random forest regression Random forest classification Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary 175 Summary 175 Chapter 7: Neural Networks and Deep Learning 177 Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning PH2O background Data upload to H2O Create train and test datasets
Random forest classification Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary 175 Summary 176 Chapter 7: Neural Networks and Deep Learning 177 Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets 163 167 172 172 173 175 176 177 177 177 177 177 178 179 179 170 179 170 179 170 170 170 170 170 170 170 170 170 170
Extreme gradient boosting - classification Model selection Feature Selection with random forests Summary Chapter 7: Neural Networks and Deep Learning Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets
Model selection Feature Selection with random forests Summary 175 Chapter 7: Neural Networks and Deep Learning 177 Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets 172 172 175 176 177 177 177 177 177 177 177 177 177
Feature Selection with random forests Summary Chapter 7: Neural Networks and Deep Learning Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding 186 Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets 172 175 176 177 177 178 179 179 179 170 179 179 179 179 179 179 179 179 179 179
Summary Chapter 7: Neural Networks and Deep Learning Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets
Chapter 7: Neural Networks and Deep Learning177Introduction to neural networks177Deep learning, a not-so-deep overview182Deep learning resources and advanced methods184Business understanding186Data understanding and preparation187Modeling and evaluation192An example of deep learning197H2O background198Data upload to H2O198Create train and test datasets200
Introduction to neural networks Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets
Deep learning, a not-so-deep overview Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets 182 184 185 186 Data understanding and preparation 187 Modeling and evaluation 198 198 200
Deep learning resources and advanced methods Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets
Business understanding Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets 186 187 187 188 198 199 200
Data understanding and preparation Modeling and evaluation An example of deep learning H2O background Data upload to H2O Create train and test datasets 187 198 200
Modeling and evaluation192An example of deep learning197H2O background198Data upload to H2O198Create train and test datasets200
An example of deep learning H2O background Data upload to H2O Create train and test datasets 197 198 200
H2O background 198 Data upload to H2O 198 Create train and test datasets 200
Data upload to H2O Create train and test datasets
Create train and test datasets 200
Summany Siller of the Country of the House of those and the Country of the Countr
Chapter 8: Cluster Analysis
Hierarchical clustering and is bremmoser bas, not suls ve, palleboll 208
Distance calculations alaylans stab Isitneupe 209
K-means clustering beilggs staylers Istineupe2 210
Gower and partitioning around medoids visminus 211
Gower 212
PAM 213
Random forest 213
Business understanding 214
Business understanding 214 Data understanding and preparation 215 Modeling and evaluation 217
Modeling and evaluation 217

	Hierarchical clustering seuplindest entitle welvieve nA	217
	K-means clustering	227
	Gower and PAM	231
	Random Forest and PAM	233
	Summary gnitecod theibard	235
Cha	Chapter 9: Principal Components Analysis	
154 1882	An overview of the principal components	238
	Rotation	241
	Business understanding noiseangen teerof mobile 8	243
	Data understanding and preparation	244
	Modeling and evaluation	246
	Component extraction	246
	Orthogonal rotation and interpretation	247
	Creating factor scores from the components	249
	Regression analysis	250
	Summary ahowies issues of neitouborist	256
Cha	pter 10: Market Basket Analysis, Recommendation Engines, and	
	uential Analysis shortem bearies and advanced methods	257
186	An overview of a market basket analysis	258
	Business understanding	259
	Data understanding and preparation	260
	Modeling and evaluation	262
	An overview of a recommendation engine	266
	User-based collaborative filtering	268
	Item-based collaborative filtering	268
	Singular value decomposition and principal components analysis	269
	Business understanding and recommendations	273
	Data understanding, preparation, and recommendations	273
	Modeling, evaluation, and recommendations	276
	Sequential data analysis	286
	Sequential analysis applied	287
	Summary ablobem bouous golinoitinsq bos newoo	294
Cha	pter 11: Creating Ensembles and Multiclass Classification	295
213	Ensembles	296
	Business and data understanding	297
	Modeling evaluation and selection	298
	Multiclass classification	301
	Business and data understanding	302

	Model evaluation and selection	Summary -	306
	Random forest		307
	Ridge regression		309
	MLR's ensemble		310
	Summary		312
Cha	pter 12: Time Series and Causality	-John Wayne	313
	Univariate time series analysis		314
	Understanding Granger causality		320
	Business understanding		321
	Data understanding and preparation		323
	Modeling and evaluation		327
	Univariate time series forecasting		327
	Examining the causality		331
	20100110910001011		331
	Vector autoregression		334
OI.			339
Cha	pter 13: Text Mining	great quotes, rie tono den	341
	Text mining framework and methods		342
	Topic models		344
	Other quantitative analyses		345
	Business understanding		347
	Data understanding and preparation		347
	Modeling and evaluation		350
	Word frequency and topic models		350
	Additional quantitative analysis		355
	Summary		364
Cha	pter 14: R on the Cloud	ist of all, I have surredue	365
	Creating an Amazon Web Services account		366
	Launch a virtual machine		368
	Start RStudio		372
	Summary when you have to share code with the		374
App	pendix A: R Fundamentals	ormost recent versions. V	375
	Getting R up-and-running		375
	Using R		381
	Data frames and matrices		386
	Creating summary statistics		388
	Installing and loading R packages		392
	Data manipulation with dplyr		393

App	Summary endix B: Sources		396 397
Inde		Ridge regression	399
312	Schooling	Summary	235
		Business understanding	
		Other quantitative analyses	