

Ameet V Joshi

Machine Learning and Artificial Intelligence



Springer

Contents

Part I Introduction

1	Introduction to AI and ML	3
1.1	Introduction	3
1.2	What Is AI	4
1.3	What Is ML	4
1.4	Organization of the Book	5
1.4.1	Introduction	5
1.4.2	Machine Learning	5
1.4.3	Building End to End Pipelines	6
1.4.4	Artificial Intelligence	6
1.4.5	Implementations	6
1.4.6	Conclusion	7
2	Essential Concepts in Artificial Intelligence and Machine Learning	9
2.1	Introduction	9
2.2	Big Data and Not-So-Big Data	9
2.2.1	What Is Big Data	9
2.2.2	Why Should We Treat Big Data Differently?	10
2.3	Types of Learning	10
2.3.1	Supervised Learning	10
2.3.2	Unsupervised Learning	11
2.3.3	Reinforcement Learning	11
2.4	Machine Learning Methods Based on Time	11
2.4.1	Static Learning	11
2.4.2	Dynamic Learning	12
2.5	Dimensionality	12
2.5.1	Curse of Dimensionality	13
2.6	Linearity and Nonlinearity	13
2.7	Occam's Razor	18
2.8	No Free Lunch Theorem	18

2.9	Law of Diminishing Returns.....	19
2.10	Early Trends in Machine Learning	19
2.10.1	Expert Systems	19
2.11	Conclusion.....	20
3	Data Understanding, Representation, and Visualization.....	21
3.1	Introduction.....	21
3.2	Understanding the Data	21
3.2.1	Understanding Entities	22
3.2.2	Understanding Attributes.....	22
3.2.3	Understanding Data Types	24
3.3	Representation and Visualization of the Data.....	24
3.3.1	Principal Component Analysis.....	24
3.3.2	Linear Discriminant Analysis	27
3.4	Conclusion.....	29
 Part II Machine Learning		
4	Linear Methods	33
4.1	Introduction.....	33
4.2	Linear and Generalized Linear Models	34
4.3	Linear Regression	34
4.3.1	Defining the Problem.....	34
4.3.2	Solving the Problem.....	35
4.4	Regularized Linear Regression	36
4.4.1	Regularization	36
4.4.2	Ridge Regression	36
4.4.3	Lasso Regression	37
4.5	Generalized Linear Models (GLM)	37
4.5.1	Logistic Regression.....	37
4.6	k -Nearest Neighbor (KNN) Algorithm.....	38
4.6.1	Definition of KNN.....	38
4.6.2	Classification and Regression	40
4.6.3	Other Variations of KNN.....	40
4.7	Conclusion.....	41
5	Perceptron and Neural Networks	43
5.1	Introduction.....	43
5.2	Perceptron	43
5.3	Multilayered Perceptron or Artificial Neural Network	44
5.3.1	Feedforward Operation.....	44
5.3.2	Nonlinear MLP or Nonlinear ANN.....	45
5.3.3	Training MLP.....	45
5.3.4	Hidden Layers	48
5.4	Radial Basis Function Networks	48
5.4.1	Interpretation of RBF Networks	49

5.5	Overfitting and Regularization.....	50
5.5.1	L1 and L2 Regularization	50
5.5.2	Dropout Regularization	51
5.6	Conclusion.....	51
6	Decision Trees	53
6.1	Introduction.....	53
6.2	Why Decision Trees?	54
6.2.1	Types of Decision Trees	54
6.3	Algorithms for Building Decision Trees	54
6.4	Regression Tree	55
6.5	Classification Tree.....	57
6.6	Decision Metrics	57
6.6.1	Misclassification Error	57
6.6.2	Gini Index	57
6.6.3	Cross-Entropy or Deviance	58
6.7	CHAID.....	58
6.7.1	CHAID Algorithm.....	59
6.8	Training Decision Tree.....	59
6.8.1	Steps.....	59
6.9	Ensemble Decision Trees	60
6.10	Bagging Ensemble Trees.....	60
6.11	Random Forest Trees.....	61
6.11.1	Decision Jungles	61
6.12	Boosted Ensemble Trees	62
6.12.1	AdaBoost.....	62
6.12.2	Gradient Boosting	62
6.13	Conclusion.....	63
7	Support Vector Machines	65
7.1	Introduction.....	65
7.2	Motivation and Scope	65
7.2.1	Extension to Multi-Class Classification	66
7.2.2	Extension for Nonlinear Case	66
7.3	Theory of SVM.....	67
7.4	Separability and Margins	69
7.4.1	Regularization and Soft Margin SVM	69
7.4.2	Use of Slack Variables.....	69
7.5	Nonlinearity and Use of Kernels	70
7.5.1	Radial Basis Function	70
7.5.2	Polynomial.....	71
7.5.3	Sigmoid	71
7.6	Risk Minimization	71
7.7	Conclusion.....	71

8	Probabilistic Models	73
8.1	Introduction	73
8.2	Discriminative Models	74
8.2.1	Maximum Likelihood Estimation	74
8.2.2	Bayesian Approach	74
8.2.3	Comparison of MLE and Bayesian Approach	76
8.3	Generative Models	78
8.3.1	Mixture Methods	79
8.3.2	Bayesian Networks	79
8.4	Some Useful Probability Distributions	79
8.4.1	Normal or Gaussian Distribution	80
8.4.2	Bernoulli Distribution	81
8.4.3	Binomial Distribution	84
8.4.4	Gamma Distribution	84
8.4.5	Poisson Distribution	85
8.5	Conclusion	89
9	Dynamic Programming and Reinforcement Learning	91
9.1	Introduction	91
9.2	Fundamental Equation of Dynamic Programming	91
9.3	Classes of Problems Under Dynamic Programming	93
9.4	Reinforcement Learning	93
9.4.1	Characteristics of Reinforcement Learning	93
9.4.2	Framework and Algorithm	94
9.5	Exploration and Exploitation	95
9.6	Examples of Reinforcement Learning Applications	95
9.7	Theory of Reinforcement Learning	96
9.7.1	Variations in Learning	97
9.8	Conclusion	98
10	Evolutionary Algorithms	99
10.1	Introduction	99
10.2	Bottleneck with Traditional Methods	99
10.3	Darwin's Theory of Evolution	100
10.4	Genetic Programming	102
10.5	Swarm Intelligence	104
10.6	Ant Colony Optimization	105
10.7	Simulated Annealing	106
10.8	Conclusion	106
11	Time Series Models	107
11.1	Introduction	107
11.2	Stationarity	108
11.3	Autoregressive and Moving Average Models	109
11.3.1	Autoregressive, or AR Process	110
11.3.2	Moving Average, or MA Process	110

11.3.3	Autoregressive Moving Average <i>ARMA</i> Process.....	111
11.4	Autoregressive Integrated Moving Average (<i>ARIMA</i>) Models	111
11.5	Hidden Markov Models (<i>HMM</i>)	112
11.5.1	Applications.....	114
11.6	Conditional Random Fields (<i>CRF</i>)	114
11.7	Conclusion.....	115
12	Deep Learning	117
12.1	Introduction.....	117
12.2	Origin of Modern Deep Learning	118
12.3	Convolutional Neural Networks (<i>CNNs</i>).....	119
12.3.1	1D Convolution	119
12.3.2	2D Convolution	120
12.3.3	Architecture of <i>CNN</i>	120
12.3.4	Training <i>CNN</i>	123
12.4	Recurrent Neural Networks (<i>RNN</i>)	123
12.4.1	Limitation of <i>RNN</i>	124
12.4.2	Long Short-Term Memory <i>RNN</i>	124
12.4.3	Advantages of <i>LSTM</i>	126
12.4.4	Current State of <i>LSTM-RNN</i>	126
12.5	Conclusion.....	126
13	Emerging Trends in Machine Learning	127
13.1	Introduction.....	127
13.2	Transfer Learning	127
13.3	Generative Adversarial Networks (<i>GANs</i>).....	128
13.4	Quantum Computation	128
13.4.1	Quantum Theory.....	129
13.4.2	Quantum Entanglement	130
13.4.3	Quantum Superposition	131
13.4.4	Computation with Quantum Particles	131
13.5	AutoML.....	131
13.6	Conclusion.....	132
14	Unsupervised Learning	133
14.1	Introduction.....	133
14.2	Clustering	134
14.2.1	k-Means Clustering.....	134
14.2.2	Improvements to k-Means Clustering	136
14.3	Component Analysis	137
14.3.1	Independent Component Analysis (<i>ICA</i>).....	137
14.4	Self Organizing Maps (<i>SOM</i>)	138
14.5	Autoencoding Neural Networks	138
14.6	Conclusion.....	140

Part III Building End to End Pipelines

15	Featurization	143
15.1	Introduction	143
15.2	UCI: Adult Salary Predictor	143
15.3	Identifying the Raw Data, Separating Information from Noise	144
15.3.1	Correlation and Causality	145
15.4	Building Feature Set	146
15.4.1	Standard Options of Feature Building	146
15.4.2	Custom Options of Feature Building	150
15.5	Handling Missing Values	151
15.6	Visualizing the Features	152
15.6.1	Numeric Features	152
15.6.2	Categorical Features	155
15.7	Conclusion	158
16	Designing and Tuning Model Pipelines	159
16.1	Introduction	159
16.2	Choosing the Technique or Algorithm	159
16.2.1	Choosing Technique for Adult Salary Classification	160
16.3	Splitting the Data	160
16.3.1	Stratified Sampling	162
16.4	Training	163
16.4.1	Tuning the Hyperparameters	163
16.5	Accuracy Measurement	164
16.6	Explainability of Features	164
16.7	Practical Considerations	164
16.7.1	Data Leakage	165
16.7.2	Coincidence and Causality	166
16.7.3	Unknown Categories	167
16.8	Conclusion	167
17	Performance Measurement	169
17.1	Introduction	169
17.2	Metrics Based on Numerical Error	170
17.2.1	Mean Absolute Error	170
17.2.2	Mean Squared Error	170
17.2.3	Root Mean Squared Error	170
17.2.4	Normalized Error	171
17.3	Metrics Based on Categorical Error	171
17.3.1	Accuracy	171
17.3.2	Precision and Recall	172
17.3.3	Receiver Operating Characteristics (ROC) Curve Analysis	173
17.4	Hypothesis Testing	174
17.4.1	Background	174

17.4.2	Steps in Hypothesis Testing	175
17.4.3	A/B Testing	176
17.5	Conclusion	176

Part IV Artificial Intelligence

18	Classification	179
18.1	Introduction	179
18.2	Examples of Real World Problems in Classification	179
18.3	Spam Email Detection	180
18.3.1	Defining Scope	180
18.3.2	Assumptions	181
18.3.3	Skew in the Data	182
18.3.4	Supervised Learning	182
18.3.5	Feature Engineering	183
18.3.6	Model Training	183
18.3.7	Iterating the Process for Optimization	183
18.4	Conclusion	184
19	Regression	185
19.1	Introduction	185
19.2	Predicting Real Estate Prices	185
19.2.1	Defining Regression Specific Problem	185
19.2.2	Gather Labelled Data	186
19.2.3	Feature Engineering	187
19.2.4	Model Selection	190
19.2.5	Model Performance	190
19.3	Other Applications of Regression	191
19.4	Conclusion	191
20	Ranking	193
20.1	Introduction	193
20.2	Measuring Ranking Performance	194
20.3	Ranking Search Results and Google's PageRank	196
20.4	Techniques Used in Ranking Systems	196
20.4.1	Keyword Identification/Extraction	196
20.5	Conclusion	198
21	Recommendations Systems	199
21.1	Introduction	199
21.2	Collaborative Filtering	200
21.2.1	Solution Approaches	201
21.3	Amazon's Personal Shopping Experience	202
21.3.1	Context Based Recommendation	202
21.3.2	Personalization Based Recommendation	203
21.4	Netflix's Streaming Video Recommendations	203
21.5	Conclusion	204

Part V Implementations

22	Azure Machine Learning	207
22.1	Introduction.....	207
22.2	Azure Machine Learning Studio.....	207
22.2.1	How to Start?	208
22.3	Building ML Pipeline Using AML Studio	210
22.3.1	Get the Data.....	210
22.3.2	Data Preprocessing	212
22.3.3	Training the Classifier Model	214
22.4	Scoring and Performance Metrics	215
22.4.1	Comparing Two Models.....	217
22.5	Conclusion.....	219
23	Open Source Machine Learning Libraries	221
23.1	Introduction.....	221
23.2	Options of Machine Learning Libraries	222
23.3	Scikit-Learn Library.....	223
23.3.1	Development Environment	223
23.3.2	Importing Data.....	224
23.3.3	Data Preprocessing	225
23.3.4	Splitting the Data Using Stratified Sampling.....	226
23.3.5	Training a Multiclass Classification Model	226
23.3.6	Computing Metrics	227
23.3.7	Using Alternate Model	227
23.4	Model Tuning and Optimization.....	228
23.4.1	Generalization	229
23.5	Comparison Between AML Studio and Scikit-Learn	229
23.6	Conclusion.....	232
24	Amazon’s Machine Learning Toolkit: Sagemaker	233
24.1	Introduction.....	233
24.2	Setting Up Sagemaker	233
24.3	Uploading Data to S3 Storage	234
24.4	Writing the Machine Learning Pipeline Using Python	235
24.5	Conclusion.....	236

Part VI Conclusion

25	Conclusion and Next Steps	247
25.1	Overview.....	247
25.2	What’s Next	248

References.....	249
------------------------	------------

Index.....	253
-------------------	------------