

Ameet V Joshi

Machine Learning and Artificial Intelligence



Springer

Machine Learning and Artificial Intelligence

Ameet V Joshi

Machine Learning and Artificial Intelligence



Springer

Ameet V Joshi
Microsoft (United States)
Redmond, WA, USA

ISBN 978-3-030-26621-9 ISBN 978-3-030-26622-6 (eBook)
<https://doi.org/10.1007/978-3-030-26622-6>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To everyone who believes in human
intelligence to be mature enough to coexist
with machine intelligence and be benefitted
and entertained by it. . .*

Foreword

I first got exposed to computing about 35 years back. It was the time when the term “microcomputers” was being replaced by the term “personal computers (PC).” Within a space of 5 years, there was this seismic shift that happened during which computing as a superspecialized domain with a total of only a few thousand people across the world working with these magical machines in cooled clean rooms suddenly became readily accessible to hundreds of millions. It dawned on everyone that computing was going to impact every aspect of business and eventually personal lives too. With that came the scramble to learn computing—whether it be about the basics of what a computer is or how to use applications like word processing or spreadsheets or engineers in every part of the world swarming to learn computer architectures and programming, algorithms, and data structures. If you were not learning any of those to some level of competency, you were heavily disadvantaged in your professions—regardless of the profession.

Fast forward 15 years to about 1995. PCs had evolved and become powerful and had taken over business computing with client-server architectures, and most people in developed countries had PCs of their own at work and home. Then, the terms Internet, web, Mosaic, Netscape, browser, webserver, HTML, and HTTP suddenly swept the computing landscape. The seismic shift that happened then entirely reframed how data and information became democratized at a scale several orders of magnitude greater than anything that preceded in the history of humanity. The Internet led to the creation of the smartphones which in turn exponentially increased the scale and reach of the Internet. The computing architectures, especially around data representation, storage, and distribution, went through a massive revolution. With that came the scramble to learn “Internet”—whether it was people of all ages learning the basics of what the Internet is, how to use a browser, and how to communicate with people electronically to authoring content as web consumables (emails, pages, sites, blogs, posts, tweets, etc.). Engineers swarmed to learn Internet architectures, web and mobile application development, distributed data architectures, etc. If you were not learning something along this spectrum, you were not only heavily disadvantaged professionally, you were probably disadvantaged socially.

Fast forward another 15 years. Advances in the Internet and mobile technologies lead to explosive growth in creation and aggregation of data. The massive aggregation of data and the virtually infinite computational capacity that goes with it evolved to being physically distributed and yet logically unified. This became the “Cloud.” Organizing and making sense of this scale of data were not only beyond human capabilities, it was even beyond the capabilities of traditional algorithms. Traditional algorithms couldn’t even scale to the “searching” through this sea of data let alone make sense of it. Yet, the easy availability of such vast volumes of data finally makes possible the long-held dream of computers transcending from “processing information” to actually “creating intelligence” by making sense of the data. Decades of research in ML and AI became re-energized and have made progress in leaps and bounds with the easy access to massive quantities of data and computation in the cloud. Thus, we are officially now in the age of “AI.” Just as has happened in the PC and the Internet revolutions, there is now a realization that AI is going to transform not just computing but society and humanity in unprecedented ways. It is beginning to dawn on people that applicability of AI and machine learning go far beyond mundane things like facial recognition in their Facebook feeds. It is also beyond the flashy, sci-fi scenarios like self-driving cars and digital assistants. Businesses of every kind—health care, manufacturing, hospitality, financial, etc.—are in the process of massive transformation with data and AI at the center of it. In all facets of life, digital information without intelligence will become pointless. The next decade will be transformed by AI. The Internet age is now set to transition to the age of AI.

However, the “swarm” towards understanding AI is just beginning. In that context, I am thrilled that Dr. Ameet Joshi is coming out with this book on Machine Learning and Artificial Intelligence. When a topic is of such broad human interest and impact, one can never have enough sources of knowledge and learning. Compared to the kind of learning content that was created in the PC era or the Internet era, machine learning and AI related content is still sparse and somewhat narrowly targeted. This is an area which is still mysterious even for the technically savvy folks. While there is a need to address the full spectrum of audiences, there is an urgent need to demystify the space for folks who are comfortable with the math and science without necessarily being computer scientists or engineers. These are folks who are willing and capable of getting a good handle on ML and AI. They are folks who understand data and math at fairly advanced levels; they possibly understand traditional computing and algorithms—but they have been left off the island populated by the “priests of data science.” They are eager to roll up their sleeves, dive in, and understand what ML and AI from its foundations. ML and AI learning is ripe for democratization. Dr. Joshi’s book hits the exact right sweet spot on this. Most people who read this may not become data scientists—but all of them will gain a perspective on how ML and AI work and will be able to take it to whatever domain of work they will be involved in. They will be able to see every application of AI, be it a digital assistant they converse with or a spookily accurate targeting of an advertisement or price fluctuations of airline ticket prices in a very different light and with a much deeper understanding. The shroud of mystery around

many things digital happening around us will seem more obvious and matter of fact. More importantly, I fully expect people who read this book to use that learning to look for applications in their future domains of work. For some set of readers, it may even intrigue them enough to become practitioners of ML and AI as a profession. There is definitely a huge demand for such professionals.

Dr. Joshi is a passionate practitioner himself. His passion shows through in this book. He has worked on broad set of ML and AI problems, and that practical experience has helped him identify and explain relatable scenarios and applications in this book. This is one of those books which likely is more than a onetime read. In all likelihood, the readers will want to revisit sections of the book to refresh their understanding or dig deeper in an area when they see real-world applications of the techniques described in this book. I think those are the best kinds of books to own. Even as an industry veteran, I enjoyed reading the book and will keep a copy handy.

Happy reading.

General Manager
Microsoft
Redmond, Washington, USA

Vij Rajarajan

Preface

One of the greatest physicists of all time and Nobel Laureate Dr. Richard Feynman was once asked by his peer to explain a property of Fermi-Dirac statistics that was very recently discovered then. Feynman, quickly said,

Not only I will explain it to you, but I will prepare a lecture on it for freshman level.

However, quite unusually, after few days, he came back and admitted,

I could not do it. I just could not reduce the explanation to freshman level. That means we really don't understand it.

It was quite a bold remark coming from even Dr. Feynman. However, apart from the topic of Fermi-Dirac statistics itself, it alludes to a very deep thought about our understanding of things in general. Freshman level here essentially meant something that can be derived directly using the first principles in mathematics or physics. This thought has always made me conscious to try and explain everything that I claim to understand using first principles, try to explain everything conceptually and not only using elaborate set of equations.

The area of *artificial intelligence* and *machine learning* has exploded over the last decade. With the widespread popularity, the core concepts in the field have been sometimes diluted and sometimes reinterpreted. With such exponential growth in the area, the scope of the field has also grown in proportion. A newcomer in the area can quickly find the topic daunting and confusing. One can always start with searching the relevant topics on the web or just start with Wikipedia, but more often than not, every single topic opens a rabbit hole with more and more new and unknown concepts, and one can get lost very easily. Also, most of the concepts in machine learning are deeply rooted in mathematics and statistics. Without solid background in theoretical mathematics and statistics, the sophisticated derivations of the theorems and lemmas can make one feel confused and disinterested in the area.

I have made an attempt here to introduce most fundamental topics in machine learning and their applications to build artificially intelligent solutions with an intuitive and conceptual approach. There would be some mathematical guidance

used from time to time, without which the concepts would not be sufficiently clear, but I have tried to avoid complex derivations and proofs to make the content more accessible to readers who are not from strong mathematical background. In the process, as per Dr. Feynman, I also make sure I have understood them myself. As far as general mathematical and statistical requirements go, I would say that typical undergraduate level should suffice. Also, with proliferation and standardization of the machine learning libraries in open-source domain, one does not need to go that deep into mathematical understanding of the theory to be able to implement the state-of-the-art machine learning models, leading to the state-of-the-art intelligent solutions.

One of the main sources of confusion that arises when trying to solve a problem in the given application is the choice of algorithm. Typically, each algorithm presented here has originated from some specific problem, but the algorithm is typically not restricted to solving only that problem. However, choosing the right algorithm for the given problem is not trivial even for a doctoral fellow with strong mathematical background. In order to separate these two areas, I have divided these two areas into separate parts altogether. This will make the topics much easier to access for the reader.

I would recommend the reader to start with *Part I* and then choose *Part II* or *Part III*, depending on the needs. It will be ideal for a student to go sequentially through the book, while a newcomer to the area from professional background would be better suited to start with *Part III* to understand or focus on the precise application at hand and then delve into the details of the theory for the algorithms as needed in *Part II*. *Part IV* and *Part V* should follow afterwards. I have added sufficient references between the two parts to make this transition smooth.

In my mind, unless one can see the models in action on real data that one can see and plot, the understanding is not complete. Hence, following the details of algorithms and applications, I have added another part to cover the basic implementation of the models using free and open-source options. Completion of this part will enable the reader to tackle the real-world problems in AI with state-of-the-art ML techniques!

Redmond, WA, USA
March 2019

Ameet Joshi

Acknowledgments

I would like to use this opportunity to acknowledge the people who had deep impact on the creation of this book. This, being the first ever book written by me, was quite a challenging task. There were quite a few times when I had almost given up. However, the continuous support and encouragement from my wife, Meghana, and sons, Dhroov and Sushaan, really helped me continue with the efforts and ultimately complete the book. I would also like to thank my father, Vijay, and brother, Mandar, for their feedback and support.

I would like to thank Mary James of Springer publications for the encouragement and support and understanding as I worked through the completion of the book.

The present book is an exercise in unification of the disparate areas in the field of machine learning to produce artificially intelligent experiences, and hence, the book essentially stands on the pillars of this knowledge that is created by the numerous extraordinary scientists and brilliant mathematicians over several decades in the past. So, I would like to thank them as well.

Last but not the least, I have used multiple quotes from some of the landmark movies produced over the last 60–70 years that depicted various scenarios of AI. These movies, like *2001: A Space Odyssey*, *Star Wars*, *WarGames*, *Terminator 2: Judgment Day*, and *The Matrix*, have showcased the power and scope of what AI can do and how it can change our lives. These cinematic interpretations express these extremely complex and technical concepts to laymen in highly influential manner and prepare them for what is coming in the future. Without them, the technological progress in this area would remain disconnected from people, so I would like to thank the creators of these movies.

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 (ARIMA) Models	111
11.5	Hidden Markov Models (HMM)	112
11.5.1	Applications.....	114
11.6	Conditional Random Fields (CRF)	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 (CNNs).....	119
12.3.1	1D Convolution	119
12.3.2	2D Convolution	120
12.3.3	Architecture of CNN	120
12.3.4	Training CNN.....	123
12.4	Recurrent Neural Networks (RNN)	123
12.4.1	Limitation of RNN	124
12.4.2	Long Short-Term Memory RNN.....	124
12.4.3	Advantages of LSTM.....	126
12.4.4	Current State of LSTM-RNN	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 (GANs).....	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 (ICA).....	137
14.4	Self Organizing Maps (SOM)	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
--------------	-----

Part I

Introduction

This is your last chance. After this, there is no turning back. You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill - you stay in Wonderland and I show you how deep the rabbit-hole goes. The question is, which pill would you choose?

—Morpheus, “The Matrix”

Part Synopsis

This part introduces the concepts of Artificial Intelligence (AI) and Machine Learning (ML) in the modern context. Various scenarios where these concepts are used will also be discussed. Going further, this part also discusses the understanding, representation and visualization aspects of data that form the foundation to the next topics. This will serve as a good starting point to delve into the details of the techniques, applications and implementations.

Chapter 1

Introduction to AI and ML



1.1 Introduction

It is an understatement to say that the field of Artificial Intelligence and Machine Learning has exploded in last decade. There are some rather valid reasons for the exponential growth and some not so much. As a result of this growth and popularity, there are also some of the most wrongly used words. Along with the enormous growth of the field it has also become one of the most confusing fields to understand the scope of. The very definitions of the terms AI and ML have also become diluted and vague.

We are not going to try and define these entities with word play, but rather are going to learn the context around the origins and scope of these entities. This will make their meaning apparent. The roots of these words originate from multiple disciplines and not just computer science. These disciplines include pure mathematics, electrical engineering, statistics, signal processing, and communications along with computer science to name the top few. I cannot imagine any other area that emerges as a conflation of such wide variety of disciplines. Along with the wide variety of origins, the field also finds applications in even greater number of industries ranging from high-tech applications like image processing, natural language processing to online shopping to nondestructive testing of nuclear power plants and so on.

In spite of being such multi-disciplinary, there are several things that we can conceptually learn collectively and obtain clarity of its scope. That is the primary objective behind this book. I want to make the field accessible to newcomers to the field in the form of graduate level students to engineers and professionals who are entering the field or even actively working in this field.

1.2 What Is AI

Alan Turing defined Artificial Intelligence as follows: “If there is a machine behind a curtain and a human is interacting with it (by whatever means, e.g. audio or via typing etc.) and if the human feels like he/she is interacting with another human, then the machine is artificially intelligent.” This is quite a unique way to define AI. It does not directly aim at the notion of intelligence, but rather focusses on human-like behavior. As a matter of fact, this objective is even broader in scope than mere intelligence. From this perspective, AI does not mean building an extraordinarily intelligent machine that can solve any problem in no time, but rather it means to build a machine that is capable of human-like behavior. However, just building machines that mimic humans does not sound very interesting. As per modern perspective, whenever we speak of AI, we mean machines that are capable of performing one or more of these tasks: understanding human language, performing mechanical tasks involving complex maneuvering, solving computer-based complex problems possibly involving large data in very short time and revert back with answers in human-like manner, etc.

The supercomputer depicted in movie 2001: A space odyssey, called HAL represents very closely to modern view of AI. It is a machine that is capable of processing large amount of data coming from various sources and generating insights and summary of it at extremely fast speed and is capable of conveying these results to humans in human-like interaction, e.g., voice conversation.

There are two aspects to AI as viewed from human-like behavior standpoint. One is where the machine is intelligent and is capable of communication with humans, but does not have any locomotive aspects. HAL is example of such AI. The other aspect involves having physical interactions with human-like locomotion capabilities, which refers to the field of robotics. For the purpose of this book, we are only going to deal with the first kind of AI.

1.3 What Is ML

The term “Machine Learning” or ML in short, was coined in 1959 by Arthur Samuel in the context of solving game of checkers by machine. The term refers to a computer program that can learn to produce a behavior that is not explicitly programmed by the author of the program. Rather it is capable of showing behavior that the author may be completely unaware of. This behavior is learned based on three factors: (1) Data that is consumed by the program, (2) A metric that quantifies the error or some form of distance between the current behavior and ideal behavior, and (3) A feedback mechanism that uses the quantified error to guide the program to produce better behavior in the subsequent events. As can be seen the second and third factors quickly make the concept abstract and emphasizes deep

mathematical roots of it. The methods in machine learning theory are essential in building artificially intelligent systems.

1.4 Organization of the Book

I have divided this book into six parts, as described below.

Parts of the Book

1. Introduction
2. Machine Learning
3. Building end to end pipelines
4. Artificial Intelligence
5. Implementations
6. Conclusion

1.4.1 Introduction

The first part of the book introduces the key concepts in the field as well as outlines the scope of the field. There are multiple aspects discussed in these chapters that may appear disconnected, but put together they are all extremely important in holistic understanding of the field and inspiring the confidence. This part also discusses the preliminary understanding and processing of the raw data as is typically observed. This sets the foundation for the reader to understand the subsequent parts.

1.4.2 Machine Learning

Second part of the book deals with the theoretical foundation of the machine learning. In this part we will study various algorithms, their origins and their applications as well. Machine learning has united a vast array of algorithms that find their origins in fields ranging from electrical engineering, signal processing, statistics, financial analysis, genetic sciences, and so on. All these algorithms are primarily developed from the principles of pure mathematics and statistics, in spite of being originated in fundamentally different areas. Along with the roots, they also have one more thing in common: the use of computers to automate complex computations. These computations ultimately lead to solving problems that seem so hard that one would believe that they are solved by some intelligent entity, or *Artificial Intelligence*.

1.4.3 Building End to End Pipelines

After learning about the theory of machine learning, one is *almost* ready to start solving problems in real world. However, the small gap that still remains is addressed in this part. Putting together theory and application is the primary objective to be learned from this part. Invariably the machine learning system built to solve a real life problem always involves multiple components being operated in sequence, creating a pipeline of operations. This part discusses all the essential components of such pipeline that takes the raw data at the input and produces interpretable results at the output.

1.4.4 Artificial Intelligence

This is likely the most interesting part of the book that directly deals with problems that machine learning has solved and we experience in our day-to-day life. These applications represent real problems that have been solved in variety of different industries, e.g., face recognition requirements for national security, detecting spam emails for e-commerce providers, deploying drones to carry out tactical jobs in areas where situations are hazardous for humans, etc. These applications need help from the rich collection of machine learning techniques that are described in the previous part. It is important to separate the applications and techniques, as there is a non-trivial overlap between those areas. There are some techniques specifically developed for certain types of applications, but there are some applications that can be benefitted from variety of different techniques, and the relationship between them is many to many. Only when we start looking at the field from these angles separately, lot of confusion starts to melt away and things start to make sense. The seamless solution of these problems and their delivery to millions of people is what creates the feel of artificial intelligence.

1.4.5 Implementations

The completion of studying parts 1 and 2, one can have sufficient confidence in analyzing a given problem in the areas of machine learning and artificial intelligence and come up with an outline of solution. However, this outline is not sufficient in actually implement the solution. In most cases, the theory is quite complex and solution of optimization problem is non-trivial. I am not going to delve into deep mathematical proofs and flowcharts that can be directly translated into computer programs for two reasons:

1. The content can quickly become too mathematical and will lose the interest of most of the target audience.

2. The complete theory behind all the techniques discussed here is so vast, that it is just not within the scope of a single book.

However, what I am going to target here is a conceptual understanding of the techniques, that is sufficient for a professional computer engineer, or a data analyst/scientist to grasp underlying key parameters of the problem.

Fortunately, the open source libraries in machine learning have matured to such level that, with such conceptual understanding of the techniques and applications, one can confidently implement reliable and robust artificially intelligent solutions without need to fully understand the underlying mathematics. This is the target of the part 3 of the book. This part will truly enable the reader to go ahead and build end to end solution of a complex practical problem and demonstrate an artificially intelligent system.

1.4.6 Conclusion

In this last part we summarize the learnings and discuss the next steps for the reader. After completion of this book the reader is well equipped to embark on solving real world problems that are still unsolved and make the world a better place.

Chapter 2

Essential Concepts in Artificial Intelligence and Machine Learning



2.1 Introduction

There are various concepts that are spread all over the place in terms of categorizing them with techniques or applications or implementations. However, they are at the heart of the entire machine learning theory and its applications and need special attention. We will cover these topics in this chapter. We will keep revising them in the book, but aggregating them here will make the reader more equipped to study the subsequent chapters.

2.2 Big Data and Not-So-Big Data

2.2.1 *What Is Big Data*

Big data processing has become a very interesting and exciting chapter in the field of AI and ML since the advent of cloud computing. Although there is no specific memory size above which the data is called big, the generally accepted definition is as follows: When the size of data is large enough such that it cannot be processed on a single machine, it is called as big data. Based on current (2018) generation of computers this equates to something roughly more than 10 GB. From there it can go to hundreds of petabytes (1 petabyte is 1000 terabytes and 1 terabyte is 1000 gigabytes) and more.

2.2.2 Why Should We Treat Big Data Differently?

Although the big data is separated from the not-so-big-data by simple size restrictions, the handling of the data changes drastically when we move from not-so-big data to big data. For processing not-so-big data, one need not worry about the location where each bit of data is stored. As all the data can be loaded into memory of single machine and can be accessed without additional overhead, all the typical numeric or string operations are executed in predictable time. However, once we enter the realm of big data, all such bets are off. Instead of a single machine, we are dealing with a cluster of machines. Each machine has its own storage and computational resources. For all the numeric and string operations one must carefully split the data into different machines, such that all the operands in individual operation are available on the same machine. If not, there is additional overhead of moving the data across the machines and it typically a very computation intensive operation. Iterative operations are especially very computation intensive for big data. Thus when dealing with big data, one must also carefully design the storage along with the operations.

2.3 Types of Learning

The machine learning algorithms are broadly classified into three types:

1. Supervised learning algorithms
2. Unsupervised learning algorithms
3. Reinforcement learning algorithms

Let's look at each type in detail.

2.3.1 Supervised Learning

For simplicity, let's consider the ML system as a black box that produces some output when given some input. If we already have a historical data that contains the set of outputs for a set of inputs, then the learning based on this data is called as supervised learning. A classic example of supervised learning is classification. Say we have measured 4 different properties (sepal length, sepal width, petal length, and petal width) of 3 different types of flowers (Setosa, Versicolor, and Virginica) [3]. We have measurements for say 25 different examples of each type of flower. This data would then serve as training data where we have the inputs (the 4 measured properties) and corresponding outputs (the type of flower) available for training the model. Then a suitable ML model can be trained in supervised manner. Once the model is trained, we can then classify any flower (between the three known types) based on the sepal and petal measurements.