Machine Learning

Algorithms and Applications

Mohssen Mohammed
Muhammad Badruddin Khan
Eihab Bashier Mohammed Bashier

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20160428

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

Library of Congress Cataloging-in-Publication Data

Names: Mohammed, Mohssen, 1982- author. | Khan, Muhammad Badruddin, author. | Bashier, Eihab Bashier Mohammed, author.
Title: Machine learning : algorithms and applications / Mohssen Mohammed, Muhammad Badruddin Khan, and Eihab Bashier Mohammed Bashier.
Description: Boca Raton : CRC Press, 2017. | Includes bibliographical references and index.
Identifiers: LCCN 2016015290 | ISBN 9781498705387 (hardcover : alk. paper)
Subjects: LCSH: Machine learning. | Computer algorithms.
Classification: LCC Q325.5 .M63 2017 | DDC 006.3/12--dc23
LC record available at https://lccn.loc.gov/2016015290

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface xiii
Acknowledgments xv
Authors xvii
Introduction xix

1 Introduction to Machine Learning **1**
1.1 Introduction 1
1.2 Preliminaries 2
1.2.1 Machine Learning: Where Several Disciplines Meet 4
1.2.2 Supervised Learning 7
1.2.3 Unsupervised Learning 9
1.2.4 Semi-Supervised Learning 10
1.2.5 Reinforcement Learning 11
1.2.6 Validation and Evaluation 11
1.3 Applications of Machine Learning Algorithms 14
1.3.1 Automatic Recognition of Handwritten Postal Codes 15
1.3.2 Computer-Aided Diagnosis 17
1.3.3 Computer Vision 19
1.3.3.1 Driverless Cars 20
1.3.3.2 Face Recognition and Security 22
1.3.4 Speech Recognition 22

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

1.3.5 Text Mining 23
1.3.5.1 Where Text and Image Data Can Be Used Together 24
1.4 The Present and the Future 25
1.4.1 Thinking Machines 25
1.4.2 Smart Machines 28
1.4.3 Deep Blue 30
1.4.4 IBM's Watson 31
1.4.5 Google Now 32
1.4.6 Apple's Siri 32
1.4.7 Microsoft's Cortana 32
1.5 Objective of This Book 33
References 34

SECTION I SUPERVISED LEARNING ALGORITHMS

2 Decision Trees 37
2.1 Introduction 37
2.2 Entropy 38
2.2.1 Example 38
2.2.2 Understanding the Concept of Number of Bits 40
2.3 Attribute Selection Measure 41
2.3.1 Information Gain of ID3 41
2.3.2 The Problem with Information Gain 44
2.4 Implementation in MATLAB® 46
2.4.1 *Gain Ratio* of C4.5 49
2.4.2 Implementation in MATLAB 51
References 52

3 Rule-Based Classifiers 53
3.1 Introduction to Rule-Based Classifiers 53
3.2 Sequential Covering Algorithm 54
3.3 Algorithm 54
3.4 Visualization 55
3.5 Ripper 55
3.5.1 Algorithm 56

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

3.5.2 Understanding Rule Growing Process 58
3.5.3 Information Gain 65
3.5.4 Pruning...66
3.5.5 Optimization 68
References 72

4 Naïve Bayesian Classification................................73
4.1 Introduction 73
4.2 Example 74
4.3 Prior Probability 75
4.4 Likelihood...75
4.5 Laplace Estimator.. 77
4.6 Posterior Probability 78
4.7 MATLAB Implementation 79
References 82

5 The *k*-Nearest Neighbors Classifiers......................83
5.1 Introduction 83
5.2 Example 84
5.3 *k*-Nearest Neighbors in MATLAB®.......................... 86
References 88

6 Neural Networks...89
6.1 Perceptron Neural Network 89
6.1.1 Perceptrons 90
6.2 MATLAB Implementation of the Perceptron Training and Testing Algorithms 94
6.3 Multilayer Perceptron Networks 96
6.4 The Backpropagation Algorithm............................. 99
6.4.1 Weights Updates in Neural Networks..........101
6.5 Neural Networks in MATLAB 102
References 105

7 Linear Discriminant Analysis 107
7.1 Introduction 107
7.2 Example 108
References 114

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

8 Support Vector Machine 115
8.1 Introduction 115
8.2 Definition of the Problem 116
8.2.1 Design of the SVM 120
8.2.2 The Case of Nonlinear Kernel..................126
8.3 The SVM in MATLAB®..127
References 128

SECTION II UNSUPERVISED LEARNING ALGORITHMS

9 *k*-Means Clustering 131
9.1 Introduction 131
9.2 Description of the Method...................................132
9.3 The *k*-Means Clustering Algorithm.......................133
9.4 The *k*-Means Clustering in MATLAB®..................134

10 Gaussian Mixture Model....................................137
10.1 Introduction..137
10.2 Learning the Concept by Example 138
References 143

11 Hidden Markov Model 145
11.1 Introduction 145
11.2 Example..146
11.3 MATLAB Code 148
References 152

12 Principal Component Analysis............................153
12.1 Introduction 153
12.2 Description of the Problem..................................154
12.3 The Idea behind the PCA 155
12.3.1 The SVD and Dimensionality Reduction 157
12.4 PCA Implementation 158
12.4.1 Number of Principal Components to Choose 159
12.4.2 Data Reconstruction Error 160

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

12.5 The Following MATLAB® Code Applies the PCA 161
12.6 Principal Component Methods in Weka 163
12.7 Example: Polymorphic Worms Detection Using PCA 167
12.7.1 Introduction 167
12.7.2 SEA, MKMP, and PCA 168
12.7.3 Overview and Motivation for Using String Matching 169
12.7.4 The KMP Algorithm 170
12.7.5 Proposed SEA 171
12.7.6 An MKMP Algorithm 173
12.7.6.1 Testing the Quality of the Generated Signature for Polymorphic Worm A 174
12.7.7 A Modified Principal Component Analysis 174
12.7.7.1 Our Contributions in the PCA 174
12.7.7.2 Testing the Quality of Generated Signature for Polymorphic Worm A 178
12.7.7.3 Clustering Method for Different Types of Polymorphic Worms 179
12.7.8 Signature Generation Algorithms Pseudo-Codes 179
12.7.8.1 Signature Generation Process 180
References 187

Appendix I: Transcript of Conversations with Chatbot 189

Appendix II: Creative Chatbot 193

Index 195

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

Introduction

Since their evolution, humans have been using many types of tools to accomplish various tasks. The creativity of the human brain led to the invention of different machines. These machines made the human life easy by enabling people to meet various life needs, including travelling, industries, constructions, and computing.

Despite rapid developments in the machine industry, intelligence has remained the fundamental difference between humans and machines in performing their tasks. A human uses his or her senses to gather information from the surrounding atmosphere; the human brain works to analyze that information and takes suitable decisions accordingly. Machines, in contrast, are not intelligent by nature. A machine does not have the ability to analyze data and take decisions. For example, a machine is not expected to understand the story of Harry Potter, jump over a hole in the street, or interact with other machines through a common language.

The era of intelligent machines started in the mid-twentieth century when Alan Turing thought whether it is possible for machines to think. Since then, the artificial intelligence (AI) branch of computer science has developed rapidly. Humans have had the dreams to create machines that have the same level of intelligence as humans. Many science fiction movies have expressed these dreams, such as *Artificial Intelligence*; *The Matrix*; *The Terminator*; *I, Robot*; and *Star Wars*.

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

The history of AI started in the year 1943 when Waren McCulloch and Walter Pitts introduced the first neural network model. Alan Turing introduced the next noticeable work in the development of the AI in 1950 when he asked his famous question: can machines think? He introduced the B-type neural networks and also the concept of test of intelligence. In 1955, Oliver Selfridge proposed the use of computers for pattern recognition.

In 1956, John McCarthy, Marvin Minsky, Nathan Rochester of IBM, and Claude Shannon organized the first summer AI conference at Dartmouth College, the United States. In the second Dartmouth conference, the term *artificial intelligence* was used for the first time. The term *cognitive science* originated in 1956, during a symposium in information science at the MIT, the United States.

Rosenblatt invented the first perceptron in 1957. Then in 1959, John McCarthy invented the LISP programming language. David Hubel and Torsten Wiesel proposed the use of neural networks for the computer vision in 1962. Joseph Weizenbaum developed the first expert system *Eliza* that could diagnose a disease from its symptoms. The National Research Council (NRC) of the United States founded the Automatic Language Processing Advisory Committee (ALPAC) in 1964 to advance the research in the natural language processing. But after many years, the two organizations terminated the research because of the high expenses and low progress.

Marvin Minsky and Seymour Papert published their book *Perceptrons* in 1969, in which they demonstrated the limitations of neural networks. As a result, organizations stopped funding research on neural networks. The period from 1969 to 1979 witnessed a growth in the research of knowledge-based systems. The developed programs Dendral and Mycin are examples of this research. In 1979, Paul Werbos proposed the first efficient neural network model with backpropagation. However, in 1986, David Rumelhart, Geoffrey Hinton, and

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

Ronald Williams discovered a method that allowed a network to learn to discriminate between nonlinear separable classes, and they named it *backpropagation*.

In 1987, Terrence Sejnowski and Charles Rosenberg developed an artificial neural network NETTalk for speech recognition. In 1987, John H. Holland and Arthur W. Burks invented an adapted computing system that is capable of learning. In fact, the development of the theory and application of genetic algorithms was inspired by the book *Adaptation in Neural and Artificial Systems*, written by Holland in 1975. In 1989, Dean Pomerleau proposed ALVINN (autonomous land vehicle in a neural network), which was a three-layer neural network designed for the task of the road following.

In the year 1997, the Deep Blue chess machine, designed by IBM, defeated Garry Kasparov, the world chess champion. In 2011, Watson, a computer developed by IBM, defeated Brad Rutter and Ken Jennings, the champions of the television game show *Jeopardy!*

The period from 1997 to the present witnessed rapid developments in reinforcement learning, natural language processing, emotional understanding, computer vision, and computer hearing.

The current research in machine learning focuses on computer vision, hearing, natural languages processing, image processing and pattern recognition, cognitive computing, knowledge representation, and so on. These research trends aim to provide machines with the abilities of gathering data through senses similar to the human senses and then processing the gathered data by using the computational intelligence tools and machine learning methods to conduct predictions and making decisions at the same level as humans.

The term *machine learning* means to enable machines to learn without programming them explicitly. There are four general machine learning methods: (1) supervised, (2) unsupervised, (3) semi-supervised, and (4) reinforcement learning methods. The objectives of machine learning are to enable

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

machines to make predictions, perform clustering, extract association rules, or make decisions from a given dataset.

This book focuses on the supervised and unsupervised machine learning techniques. We provide a set of MATLAB programs to implement the various algorithms that are discussed in the chapters.

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

Chapter 1

Introduction to Machine Learning

1.1 Introduction

Learning is a very personalized phenomenon for us. Will Durant in his famous book, *The Pleasures of Philosophy*, wondered in the chapter titled "Is Man a Machine?" when he wrote such classical lines:

> Here is a child; … See it raising itself for the first time, fearfully and bravely, to a vertical dignity; why should it long so to stand and walk? Why should it tremble with perpetual curiosity, with perilous and insatiable ambition, touching and tasting, watching and listening, manipulating and experimenting, observing and pondering, *growing*—till it weighs the earth and charts and measures the stars?… [1]

Nevertheless, learning is not limited to humans only. Even the simplest of species such as amoeba and paramecium exhibit this phenomenon. Plants also show intelligent

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

behavior. Only nonliving things are the natural stuffs that are not involved in learning. Hence, it seems that *living* and *learning* go together. In nature-made nonliving things, there is hardly anything to learn. Can we introduce learning in human-made nonliving things that are called *machines*? Enabling a machine capable of learning like humans is a dream, the fulfillment of which can lead us to having *deterministic machines* with *freedom* (or illusion of freedom in a sense). During that time, we will be able to happily boast that our humanoids resemble the image and likeliness of *humans* in the guise of machines.

1.2 Preliminaries

Machines are by nature not intelligent. Initially, machines were designed to perform specific tasks, such as running on the railway, controlling the traffic flow, digging deep holes, traveling into the space, and shooting at moving objects. Machines do their tasks much faster with a higher level of precision compared to humans. They have made our lives easy and smooth.

The fundamental difference between humans and machines in performing their work is intelligence. The human brain receives data gathered by the five senses: vision, hearing, smell, taste, and tactility. These gathered data are sent to the human brain via the neural system for perception and taking action. In the perception process, the data is organized, recognized by comparing it to previous experiences that were stored in the memory, and interpreted. Accordingly, the brain takes the decision and directs the body parts to react against that action. At the end of the experience, it might be stored in the memory for future benefits.

A machine cannot deal with the gathered data in an intelligent way. It does not have the ability to analyze data for

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

classification, benefit from previous experiences, and store the new experiences to the memory units; that is, machines do not learn from experience.

Although machines are expected to do mechanical jobs much faster than humans, it is not expected from a machine to: understand the play *Romeo and Juliet*, jump over a hole in the street, form friendships, interact with other machines through a common language, recognize dangers and the ways to avoid them, decide about a disease from its symptoms and laboratory tests, recognize the face of the criminal, and so on. The challenge is to make *dumb* machines learn to cope correctly with such situations. Because machines have been originally created to help humans in their daily lives, it is necessary for the machines to *think*, *understand* to solve problems, and *take* suitable decisions akin to humans. In other words, we need *smart* machines. In fact, the term *smart machine* is symbolic to machine learning success stories and its future targets. We will discuss the issue of smart machines in Section 1.4.

The question of whether a machine can think was first asked by the British mathematician Alan Turing in 1955, which was the start of the artificial intelligence history. He was the one who proposed a *test* to measure the performance of a machine in terms of intelligence. Section 1.4 also discusses the progress that has been achieved in determining whether our machines can pass the Turing test.

Computers are machines that follow programming instructions to accomplish the required tasks and help us in solving problems. Our brain is similar to a CPU that solves problems for us. Suppose that we want to find the smallest number in a list of unordered numbers. We can perform this job easily. Different persons can have different methods to do the same job. In other words, different persons can use different *algorithms* to perform the same task. These methods or algorithms are basically a sequence of instructions

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

that are executed to reach from one state to another in order to produce output from input.

If there are different algorithms that can perform the same task, then one is right in questioning which algorithm is better. For example, if two programs are made based on two different algorithms to find the smallest number in an unordered list, then for the same list of unordered number (or same set of input) and on the same machine, one measure of efficiency can be speed or quickness of program and another can be minimum memory usage. Thus, time and space are the usual measures to test the efficiency of an algorithm. In some situations, time and space can be interrelated, that is, the reduction in memory usage leading to fast execution of the algorithm. For example, an efficient algorithm enabling a program to handle full input data in cache memory will also consequently allow faster execution of program.

1.2.1 Machine Learning: Where Several Disciplines Meet

Machine learning is a branch of *artificial intelligence* that aims at enabling machines to perform their jobs skillfully by using intelligent software. The statistical learning methods constitute the backbone of intelligent software that is used to develop machine intelligence. Because machine learning algorithms require data to learn, the discipline must have connection with the discipline of database. Similarly, there are familiar terms such as Knowledge Discovery from Data (KDD), data mining, and pattern recognition. One wonders how to view the big picture in which such connection is illustrated.

SAS Institute Inc., North Carolina, is a developer of the famous analytical software Statistical Analysis System (SAS). In order to show the connection of the discipline of machine learning with different related disciplines, we will use the illustration from SAS. This illustration was actually used in a data mining course that was offered by SAS in 1998 (see Figure 1.1).

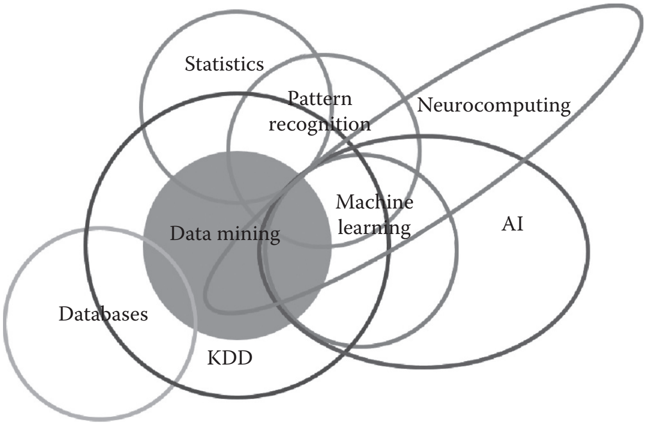

Figure 1.1 Different disciplines of knowledge and the discipline of machine learning. (From Guthrie, Looking backwards, looking forwards: SAS, data mining and machine learning, 2014, http://blogs.sas.com/content/subconsciousmusings/2014/08/22/looking-backwards-looking-forwards-sas-data-mining-and-machine-learning/2014. With permission.)

In a 2006 article entitled "The Discipline of Machine Learning," Professor Tom Mitchell [3, p.1] defined the discipline of machine learning in these words:

> Machine Learning is a natural outgrowth of the **intersection of Computer Science and Statistics**. We might say the defining question of Computer Science is 'How can we build machines that solve problems, and which problems are inherently tractable/intractable?' The question that largely defines Statistics is 'What can be inferred from data plus a set of modeling assumptions, with what reliability?' The defining question for Machine Learning builds on **both**, but it is a distinct question. Whereas Computer Science has focused primarily on how to manually program computers, Machine Learning focuses on the question of **how to get computers to program themselves (from experience plus some initial structure)**. Whereas Statistics

> has focused primarily on what conclusions can be inferred from data, Machine Learning incorporates additional questions about what computational architectures and algorithms can be used to most effectively capture, store, index, retrieve and merge these data, how multiple learning subtasks can be orchestrated in a larger system, and questions of computational tractability [emphasis added].

There are some tasks that humans perform effortlessly or with some efforts, but we are unable to explain how we perform them. For example, we can recognize the speech of our friends without much difficulty. If we are asked how we recognize the voices, the answer is very difficult for us to explain. Because of the lack of understanding of such phenomenon (speech recognition in this case), we cannot craft algorithms for such scenarios. *Machine learning* algorithms are helpful in bridging this gap of understanding.

The idea is very simple. We are not targeting to understand the underlying processes that help us learn. We write computer programs that will make machines learn and enable them to perform tasks, such as prediction. The goal of learning is to construct a *model* that takes the input and produces the desired result. Sometimes, we can understand the model, whereas, at other times, it can also be like a black box for us, the working of which cannot be intuitively explained. The model can be considered as an *approximation* of the process we want machines to mimic. In such a situation, it is possible that we obtain errors for some input, but most of the time, the model provides correct answers. Hence, another measure of performance (besides performance of metrics of speed and memory usage) of a machine learning algorithm will be the *accuracy* of results. It seems appropriate here to quote another statement about learning of computer program from Professor Tom Mitchell from Carnegie Mellon University [4, p.2]:

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

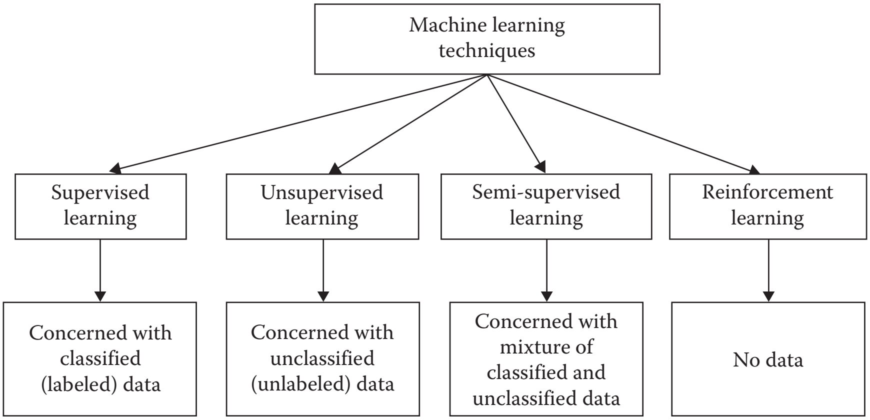

Figure 1.2 Different machine learning techniques and their required data.

> A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.

The subject will be further clarified when the issue will be discussed with examples at their relevant places. However, before the discussion, a few widely used terminologies in the machine learning or data mining community will be discussed as a prerequisite to appreciate the examples of machine learning applications. Figure 1.2 depicts four machine learning techniques and describes briefly the nature of data they require. The four techniques are discussed in Sections 1.2.2 through 1.2.5.

1.2.2 Supervised Learning

In supervised learning, the target is to infer a function or mapping from training data that is *labeled*. The training data consist of input vector **X** and output vector **Y** of labels or tags. A label or tag from vector **Y** is the *explanation* of its respective input example from input vector **X**. Together they form a *training example*. In other words, training data comprises training examples. If the labeling does not exist for input vector **X**, then **X** is *unlabeled data*.

Why such learning is called *supervised learning*? The output vector **Y** consists of labels for each training example present in the training data. These labels for output vector are provided by the supervisor. Often, these supervisors are humans, but machines can also be used for such labeling. Human judgments are more expensive than machines, but the higher error rates in data labeled by machines suggest superiority of human judgment. The manually labeled data is a precious and reliable resource for supervised learning. However, in some cases, machines can be used for reliable labeling.

Example

Table 1.1 demonstrates five unlabeled data examples that can be labeled based on different criteria.

The second column of the table titled, "Example judgment for labeling" expresses possible criterion for each data example. The third column describes possible labels after the application of judgment. The fourth column informs which actor can take the role of the supervisor.

In all first four cases described in Table 1.1, machines can be used, but their low accuracy rates make their usage questionable. Sentiment analysis, image recognition, and speech detection technologies have made progress in past three decades but there is still a lot of room for improvement before we can equate them with humans' performance. In the fifth case of tumor detection, even normal humans cannot label the X-ray data, and expensive experts' services are required for such labeling.

Two groups or categories of algorithms come under the umbrella of supervised learning. They are

1. Regression
2. Classification

Table 1.1 Unlabeled Data Examples along with Labeling Issues

Unlabeled Data Example	*Example Judgment for Labeling*	*Possible Labels*	*Possible Supervisor*
Tweet	Sentiment of the tweet	*Positive/ negative*	Human/ machine
Photo	Contains *house* and *car*	*Yes/No*	Human/ machine
Audio recording	The word *football* is uttered	*Yes/No*	Human/ machine
Video	Are weapons used in the video?	*Violent/ nonviolent*	Human/ machine
X-ray	Tumor presence in X-ray	*Present/ absent*	Experts/ machine

1.2.3 Unsupervised Learning

In unsupervised learning, we lack *supervisors* or training data. In other words, all what we have is unlabeled data. The idea is to find a hidden structure in this data. There can be a number of reasons for the data not having a label. It can be due to unavailability of funds to pay for manual labeling or the inherent nature of the data itself. With numerous data collection devices, now data is collected at an unprecedented rate. The variety, velocity, and the volume are the dimensions in which *Big Data* is seen and judged. To get something from this data without the supervisor is important. This is the challenge for today's machine learning practitioner.

The situation faced by a machine learning practitioner is somehow similar to the scene described in *Alice's Adventures in Wonderland* [5, p.100], an 1865 novel, when Alice looking to go *somewhere*, talks to the Cheshire cat.

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

> … She went on. "Would you tell me, please, **which way** I ought to go from here?"
>
> "That depends a good deal on **where** you want to get to," said the Cat.
>
> "I don't much care **where—**" said Alice.
>
> "Then it **doesn't matter** which way you go," said the Cat.
>
> "—so long as I get ***somewhere***," Alice added as an explanation.
>
> "Oh, you're sure to do that," said the Cat, "if you only **walk long enough**."

In the machine learning community, probably *clustering* (an unsupervised learning algorithm) is analogous to the *walk long enough* instruction of the Cheshire cat. The *somewhere* of Alice is equivalent to *finding regularities in the input*.

1.2.4 Semi-Supervised Learning

In this type of learning, the given data are a mixture of classified and unclassified data. This combination of labeled and unlabeled data is used to generate an appropriate model for the classification of data. In most of the situations, labeled data is scarce and unlabeled data is in abundance (as discussed previously in unsupervised learning description). The target of semi-supervised classification is to learn a model that will predict classes of future test data better than that from the model generated by using the labeled data alone. The way we learn is similar to the process of semi-supervised learning. A child is supplied with

1. Unlabeled data provided by the environment. The surroundings of a child are full of unlabeled data in the beginning.

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

2. Labeled data from the supervisor. For example, a father teaches his children about the names (labels) of objects by pointing toward them and uttering their names.

Semi-supervised learning will not be discussed further in the book.

1.2.5 Reinforcement Learning

The reinforcement learning method aims at using observations gathered from the interaction with the environment to take actions that would maximize the reward or minimize the risk.

In order to produce intelligent programs (also called *agents*), reinforcement learning goes through the following steps:

1. Input state is observed by the agent.
2. Decision making function is used to make the agent perform an action.
3. After the action is performed, the agent receives reward or reinforcement from the environment.
4. The state-action pair information about the reward is stored.

Using the stored information, policy for particular state in terms of action can be fine-tuned, thus helping in optimal decision making for our agent.

Reinforcement learning will not be discussed further in this book.

1.2.6 Validation and Evaluation

Assessing whether the model learnt from machine learning algorithm is good or not, needs both validation and evaluation. However, before discussing these two important terminologies, it is interesting to mention the writings of Plato

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

(the great philosopher) regarding this issue. The excerpt from his approach is given in Box 1.1 to introduce readers to this interesting debate.

BOX 1.1 PLATO ON STABILITY OF BELIEF

Plato's ethics is written by Terence Irwin, professor of the history of philosophy at the University of Oxford. In section 101 titled "Knowledge, Belief and Stability," there is an interesting debate about *wandering* of beliefs. The following are excerpts from the book.

> Plato also needs to consider the different circumstance that might cause true beliefs to wander away … Different demands for stability might rest on different standards of reliability. If, for instance, I believe that these animals are sheep, and they are sheep, then my belief is reliable for these animals, and it does not matter if I do not know what makes them sheep. If, however, I cannot tell the difference between sheep and goat and do not know why these animals are sheep rather than goats, my ignorance would make a difference if I were confronted with goats. If we are concerned about 'empirical reliability' (the tendency to be right in empirically likely conditions), my belief that animals with a certain appearance are sheep may be perfectly reliable (if I can be expected not to meet any goats). If we are concerned about 'counterfactual reliability' (the tendency to be right in counterfactual, and not necessarily empirically likely, conditions), my inability to distinguish sheep from goats make

(*Continued*)

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

BOX 1.1 (CONTINUED) PLATO ON STABILITY OF BELIEF

my belief unreliable that animals with certain appearance are sheep. In saying that my belief about sheep is counterfactually unreliable, we point out that my reason for believing that these things are sheep is mistaken, even though the mistake makes no difference to my judgements in actual circumstances.

When Plato speaks of a given belief 'wandering', he describes a fault that we might more easily recognize if it were described differently. If I identify sheep by features that do not distinguish them from goats, then I rely on false principles to reach the true belief 'this is a sheep' in an environment without goats. If I rely on the same principles to identify sheep in an environment that includes goats, I will often reach the false belief 'this is a sheep 'when I meet a goat. We may want to describe these facts by speaking of three things: (1) the true token belief 'this is a sheep' (applied to a sheep in the first environment), (2) the false token belief 'this is a sheep' (applied to a goat in the second environment), and (3) the false general principle that I use to identify sheep in both environments.

If one claims that for a particular training data the function fits perfectly, then for the machine learning community, this claim is not enough. They will immediately ask about the performance of function on testing data.

A function fitting perfectly on training data needs to be examined. Sometimes, it is the phenomenon of *overfitting* that will give best performance on training data, and when

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

yet-unseen labeled data will be used to test them, they will fail miserably. To avoid overfitting, it is common practice to divide the labeled data into two parts:

1. Training data
2. Testing data

A training set is used to build the model and testing set is used to validate the built model. In hold out testing/validation, one is expected to hold out part of the data for testing. Larger portion of the data is used for model training purpose, and the test metrics of the model are tested on holdout data.

The technique of cross-validation is useful when the available training dataset is quite small and one cannot afford to hold out part of the data just for validation purposes. In k-fold cross-validation, the available dataset is divided into k equal folds. Each of these k folds are treated as holdout datasets, and the training of the model is performed on rest of the $k - 1$ folds. The performance of the model is judged on the basis of holdout fold. The average of performance on all k folds is the overall performance of model.

1.3 Applications of Machine Learning Algorithms

Machine learning has proven itself to be the answer to many real-world challenges, but there are still a number of problems for which machine learning breakthrough is required. The need was felt by the cofounder and ex-chairman of Microsoft, Bill Gates, and was translated into the following wordings on one occasion [6]:

> A breakthrough in machine learning would be worth ten Microsofts

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

In this section, we will discuss some applications of machine learning with some examples.

1.3.1 Automatic Recognition of Handwritten Postal Codes

Today, in order to communicate, we use a variety of digital devices. However, the postal services still exist, helping us send our mails, gifts, and important documents to the required destination. The way machine learning has benefited this sector can be understood by citing the example of the US Postal Service.

The US Postal Service was able to exploit the potentials of machine learning in the 1960s when they successfully used machines to automatically read the city/state/ZIP code line of typed addresses to sort letters. Optical character recognition (OCR) technology was able to correctly interpret the postal address using machine learning algorithm. According to the author:

> The images that consist of typed, handwritten or printed content of text are readable for humans. In order to make such text content readable for machines, Optical character recognition technology is used.
>
> A scanned text document in an image format like bitmap is nothing but a picture of the text. OCR software analyzes the image and attempts to identify alphabetic letter and numeric digit. When it is successful in recognizing a character, it converts into machine-encoded text. This machine-encoded text can be electronically edited, searched,compressed and can be used as input for applications like automatic translation, text-to-speech and text mining. In the presence of accurate OCR, data entry becomes simple, faster and cheaper.

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

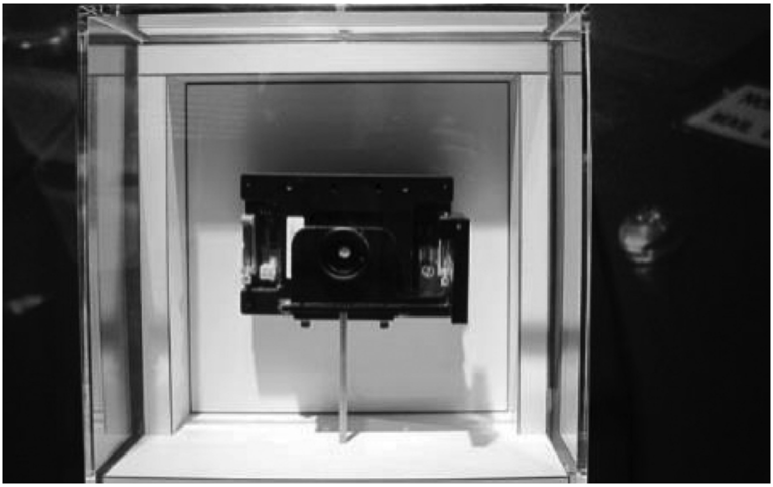

Figure 1.3 Image of OCR camera used by the US Postal Service.

In 1900, the post office was handling 7.1 billion pieces of mail per year.[*] All these were being done without cars or sophisticated machinery.

In 2006,[†] the US Postal Service sorted and delivered more than 213 billion pieces of mail, about 40% of the world's total mail volume and more than any other postal administration in the world. This enormous herculean service is provided by the US Postal Service with the help of machines.

OCR has been successful in bringing a new revolution in the efficiency of postal system. The OCR camera similar to the one shown in Figure 1.3 helped in forming the connection between the physical mail and the information system that directs it to its destination. Now improved OCR technology accompanied with other mail processing services is able to enhance the efficiency of different countries' postal services. According to the US Postal Service

[*] http://www.theatlantic.com/technology/archive/2011/12/tech-has-saved-the-postal-service-for-200-years-today-it-wont/249946/#slide10.

[†] https://about.usps.com/publications/pub100/pub100_055.htm.

website [7], "the Postal Service is the world leader in optical character recognition technology with machines reading nearly 98% of all hand-addressed letter mail and 99.5% of machine-printed mail."

Google is now providing free service to convert the text image to text documents for 200 languages in more than 25 writing systems. Google accomplished this by using hidden Markov models and treated the input as a whole sequence, rather than first trying to break it apart into pieces. The list of supported languages can be found at the Google website.* One can imagine the complexities involved in such a task. A simple problem is language identification. There no longer exists a hidden supposition that the language of the document to be processed is already known. If the language is identified wrongly, it means that we should expect a poor performance from the OCR technology.

The OCR technology is one of the applications of pattern recognition, a branch of machine learning. The focus of pattern recognition is to recognize pattern and regularities in data. The data can be text, speech, and/or image. The OCR example is the one in which input data is in the form of an image. Another example of the application of pattern recognition using image data is computer-aided diagnosis. We will discuss some of its applications in Section 1.3.2.

1.3.2 Computer-Aided Diagnosis

Pattern recognition algorithms used in computer-aided diagnosis can assist doctors in interpreting medical images in a relatively short period. Medical images from different medical tests such as X-rays, MRI, and ultrasound are the sources of data describing a patient's condition.

---

* https://support.google.com/drive/answer/176692?hl=en.

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

The responsibility of a radiologist is to analyze and evaluate the output of these medical tests that are in the form of a digital image. The short time constraint requires that the radiologist be assisted by machine. Computer-aided diagnosis uses pattern recognition techniques from machine learning to identify suspicious structures in the image. How does an algorithm catch suspicious structure? Supervised learning is done to perform this task. Few thousand labeled images are given to the machine learning algorithm, such as Bayesian classifier, artificial neural network, radial basis function network, and support vector machine. The resulting classifier is expected to classify new medical images correctly.

Mistakes in diagnosis by the machine learning algorithm can bring disaster for a family. The fault can cause damage to a person in monetary terms and it can risk his/her life, too.

The following are two such examples:

1. Suppose our classifier detects breast cancer in a patient who actually had no such disease. The results obtained by the classifier will create harmful psychological conditions for the patient. In order to confirm the result of the classifier, further tests can result in monetary losses for the patient.
2. Suppose our classifier does not detect breast cancer in patient who actually has such a disease. This will lead to wrong medical treatment and can threaten the life of the patient in near or far future. In order to avoid such mistakes, the complete substitution of doctor with technology is not recommended. The role of technology should be supportive. It should be the doctor(generally a radiologist) who must take the responsibility of the final interpretation of medical image.

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

Computer-aided diagnosis is assisting medical doctors/radiologists in the diagnosis of a number of health problems. Few examples are as follows:

- Pathological brain detection
- Breast cancer
- Lung cancer
- Colon cancer
- Prostate cancer
- Bone metastases
- Coronary artery disease
- Congenital heart defect
- Alzheimer's disease

1.3.3 Computer Vision

We want our robots to see and act appropriately after understanding the situation. The cameras installed in a robot can provide images, but they will not help the robot recognize or interpret the image. Using pattern recognition, what type of learning can a robot perform? We begin with the discussion of the example of the event called *RoboCup*.

RoboCup: "Robot Soccer World Cup" or RoboCup is an international robotic tournament of soccer. The officially declared goal of the project is very challenging and is stated as follows:

"By the middle of the 21st century, a team of fully autonomous humanoid robot soccer players shall win a soccer game, complying with the official rules of FIFA, against the winner of the most recent World Cup."* The RoboCup 2015 (China) attracted 175 intelligent sporting robot teams from 47 countries. In the largest adult size category of the event, the US team designed

* http://www.robocup.org/about-robocup/objective/.

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

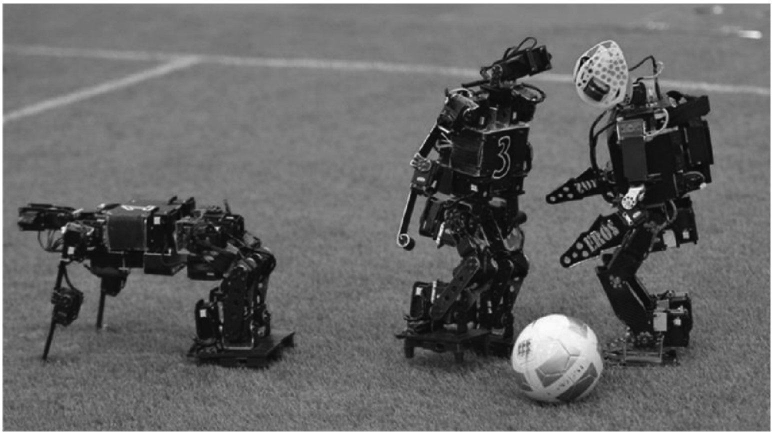

Figure 1.4 US and Iranian robot teams competing for RoboCup 2014 final. (Courtesy of Reuters.)

by the University of Pennsylvania defeated the Iranian team with tough 5–4 goal results (Figure 1.4).

The autonomous robots are expected to cooperate with their other team members (that are also robots) in adversarial dynamic environment in order to win the match. They need to categorize objects and recognize activities. To perform these tasks, they get input from their cameras. These tasks lie purely in the pattern recognition domain, a branch of machine learning.

1.3.3.1 Driverless Cars

Autonomous cars with no drivers are also one of the applications where *car vision* is actually made possible by advancement in the computer vision technology. In the industry, it is clear that there is ongoing competition to manufacture driverless cars running on the roads as early as possible. According to the BBC* report titled *Toyota promises driverless cars on roads by 2020*, different competitors

* http://www.bbc.com/news/technology-34464450.

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

are on the bandwagon and announcing their targets for driverless cars. The article states:

> **Toyota** is the latest car company to push forward with plans for an autonomous vehicle, offering fresh competition to Silicon Valley companies such as **Google**, **Cruise** and **Tesla**.
>
> Last week, **General Motors** said it was offering driverless rides to workers at its research and development facility in Warren, Michigan.
>
> **Nissan** has promised to put an automated car on Japan's roads as early as 2016.
>
> However, **Google** is already testing its self-driving cars on US city streets. And **Tesla** chief executive Elon Musk said in July his company was "almost ready" to make its cars go driverless on main roads and parallel-park themselves.

How these cars will accomplish this task? BBC article states the narrative of Toyota in following words:

> According to Toyota, the car "uses multiple external sensors to **recognise** nearby vehicles and hazards, and selects appropriate routes and lanes depending on the destination."

Based on these data inputs, it "'automatically operates the steering wheel, accelerator and brakes' to drive in much the same way as a person would drive." (Figure 1.5) [8].

The applications that are and will be using computer-vision-related technologies are very sensitive in nature. A driverless car accident can result in a tragedy for family or families. Similarly, another very sensitive area is the usage of computer vision technology in drones. The drones that are used in warfare can kill innocent people if algorithms behind the vision misbehave.

Figure 1.5 Toyota tested its self-driving Highway Teammate car on a public road. (Courtesy of BBC.)

1.3.3.2 Face Recognition and Security

Images from smart phones and CCTV cameras are now produced at an unprecedented rate. A problem pertinent to face recognition is to associate the face image to its respective identity. Building a classifier for this task is not a trivial job, because there are too many classes involved with multiple image-related problems. Face recognition can help security agencies to use a large amount of data from different sources to automatically find what is very difficult for humans to do manually.

1.3.4 Speech Recognition

The field of speech recognition aims to develop methodologies and technologies that enable computers to recognize and translate spoken language into text. Stenography (writing in shorthand) is no longer required. Automatic transcription of speech into text has found its way in areas such as video captioning and court reporting. This technology can help people with disabilities. With the passage of time, the accuracy of speech recognition engines is increasing. There is no doubt that voice-controlled programs such as Apple's Siri, Google

Now, Amazon's Alexa, and Microsoft's Cortana do not always understand our speech, but things are likely to be improved in the near future.

1.3.5 Text Mining

The examples that we have studied up until now are basically using image or voice data for learning. We have another source of learning, that is, text data. It was observed that most of the enterprise-related information is stored in text format. The challenge was how to use this unstructured data or text. The earliest definition or function of business intelligence system given by H.P. Luhn [9] in the IBM journal is as follows:

> … utilize data-processing machines for **auto-abstracting** and **auto-encoding** of documents and for creating interest profiles for each of the 'action points' in an organization. Both incoming and internally generated documents are automatically abstracted, characterized by a word pattern, and sent automatically to appropriate action points.

Another venue where the unstructured data or text is available in abundance for researchers is social media. Social media is the place where we can see the production of text data at an unprecedented level. The sharing of personal experiences in the form of text has provided stakeholders, such as business, the opportunity to analyze and use them for beneficial purpose.

Text mining is helpful in a number of applications including

- Business intelligence
- National security
- Life sciences
- Those related to sentiment classification
- Automated placement of advertisement

- Automated classification of news articles
- Social media monitoring
- Spam filter

1.3.5.1 Where Text and Image Data Can Be Used Together

It is possible that in order to solve a particular problem, both text and image data are used. For example, the problem of author identification for a particular written corpus of data can be solved in two ways:

1. *Handwriting detection*: The known corpus of handwritten data can be used to make a classifier that can assign a document to an author based on different features.
2. *Writing style detection*: This is a text mining problem. We want to find features that are related to a peculiar author using known documents attributed to the author. These features can be used to build a classifier that can identify whether the particular document belongs to the author or not.

It is possible that the two classifiers are joined together to develop a new classifier with improved performance for author identification.

Another area where such data can be helpful in solving the problem is in the identification of unwanted material in a video. In order to identify unwanted material, we can approach the problem in two ways:

1. Use video images and apply machine learning techniques on image data to make a model to identify unwanted material in the video.

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

2. Use comments from social media related to video to understand the content of the video by making a model that can predict the presence or absence of unwanted material in the video.

Once again, the two classifiers can be combined to improve the performance of the system.

1.4 The Present and the Future

1.4.1 Thinking Machines

The question of whether a machine can think was first asked by the British mathematician Alan Turing in 1955, which was the start of the artificial intelligence history. He was the one who proposed a *test* to measure the machine's performance in terms of intelligence.

In 2014, a chatbot was able to pass this Turing test (see Box 1.2 for further details). A chatbot is a computer program that simulates an intelligent conversation with one or more human users. This conversation can be performed via audio or text communication methods. Box 1.3 describes another interesting event in which one of the judges of the annual Loebner Prize* 2015 discusses the deficiencies of chatbots. We have included a full transcript of the chat between one of the judges and the 2015 winner chatbot of Loebner Prize in the Appendix I. The transcript will help readers understand how chatbots try to dodge the judges when faced with difficult questions.

Researchers at Google have programmed an advanced type of chatbot that is able to learn from training data comprising of examples from dialogues. The two sources of training data were IT helpdesk troubleshooting dataset and movie transcript dataset.

* https://en.wikipedia.org/wiki/Loebner_Prize (accessed on September 20, 2015).

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

BOX 1.2 TURING TEST PASSED BY CHATBOT NAMED EUGENE

The Turing test is based on twentieth century-mathematician and code-breaker Alan Turing's 1950 famous question and answer game, *Can Machines Think*? The experiment investigates whether people can detect if they are talking to machines or humans. If a computer is mistaken for a human more than 30% of the time during a series of a five-minute keyboard conversation, the machine passes the test.

In 2014, a computer program Eugene Goostman passed the Turing test for the first time during *Turing Test 2014* held at the renowned Royal Society in London on June 7, 2014. Eugene managed to convince 33% of the human judges (30 judges took part) that it was human.

Source: http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx

BOX 1.3 THE DIFFERENCE BETWEEN CONVERSATION WITH HUMAN AND A MACHINE

In the Tech section of the BBC website, the story appeared with the title of *AI bots try to fool human judges*, describing the live reporting of annual Loebner Prize 2015. One of the judges of the event, who had to evaluate the intelligence of a chatbot, was BBC technology correspondent Rory Cellan-Jones. The full transcript of his conversation with the 2015 prize winner, the Chatbot Rose, is given on the BBC website. The comments from Rory after the whole experience are as follows:

(*Continued*)

BOX 1.3 (CONTINUED) THE DIFFERENCE BETWEEN CONVERSATION WITH HUMAN AND A MACHINE

I feel as though I'm a psychiatrist who has just spent two hours delving into the innermost thoughts of four pairs of patients.

Being a judge in the Loebner Prize has made me think about how conversations work—and what it means to be a human conversationalist.

I quickly latched on to a simple technique for spotting the bot—be a messy human chatter.

The bots could cope with simple questions—where do you live, what do you do, how did you get here.

But the minute I started musing on London house prices, how to deal with slugs in your garden, they just fell apart.

Their technique was to try to take the conversation in another direction, ignoring what I was saying.

So, it took me no more than two or three questions to work out which was the bot and which the human.

My conclusion—it will take some time before a computer passes the Turing Test. The humans are just much more interesting to talk to.

Source: http://www.bbc.com/news/live/technology-34281198

They trained their chatbot with language model based on recurrent neural network. It means that these are not just canned answers that are given by chatbots seeing some patterns in human chats. Some of the interesting and artistic answers by

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

the chatbot from Google are available in the research paper* titled, "A neural conversational model" [10]. The researchers admitted the limitation of the work in their research paper that the chatbot was unable to have a realistic conversation currently and, hence passed the Turing test; however proper answers to many different types of questions without rules is a surprising discovery. We have included different conversations of this learning chatbot in the Appendix II.

1.4.2 Smart Machines

The dream of machines appearing as smart as humans is still far from being realized. In general, a smart machine is an *intelligent* system that uses equipment such as sensors, RFID, a Wi-Fi, or cellular communications link to receive data and interpret it to make decisions. They use machine learning algorithms to accomplish tasks usually performed by humans in an order to enhance efficiency and productivity.

Gartner, Inc.†, Stanford, California, is an American information technology (IT) research and advisory firm providing technology-related insight targeting CIOs and senior IT leaders by disseminating their research in a number of ways such as Gartner symposiums. Gartner symposium/ITxpo attracts thousands of CIOs from the industry. Gartner's analyst Kenneth F. Brant has given two criteria for a true smart machine.

A true smart machine meets two criteria‡:

1. First, a smart machine does something that no machine was ever thought to be able to do. Using that yardstick, a drone delivering a package—a model being

---

* http://arxiv.org/pdf/1506.05869v2.pdf.

† http://www.gartner.com/technology/about.jsp.

‡ http://www.forbes.com/sites/emc/2014/01/09/smart-machines-shaping-the-workforce-of-the-future/.

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

contemplated by Amazon—would qualify as a smart machine
2. Machine is capable of learning. Using the second criterion for a true smart—the delivery drone fails the test

Yet that same delivery drone–regardless of how smart it is—could still have a significant effect on productivity and employment in the shipping industry.

Smart machines were one of the top 10 technologies and trends that were predicted to be strategic for most organizations in 2014 as well as 2015 by Gartner, Inc. The prediction for 2014 placed *smart machines* in the category of *future disruption* along with the technology of 3D printers. Smart machines were again present in the prediction for 2015 in the category of *intelligence everywhere* (Figure 1.6).

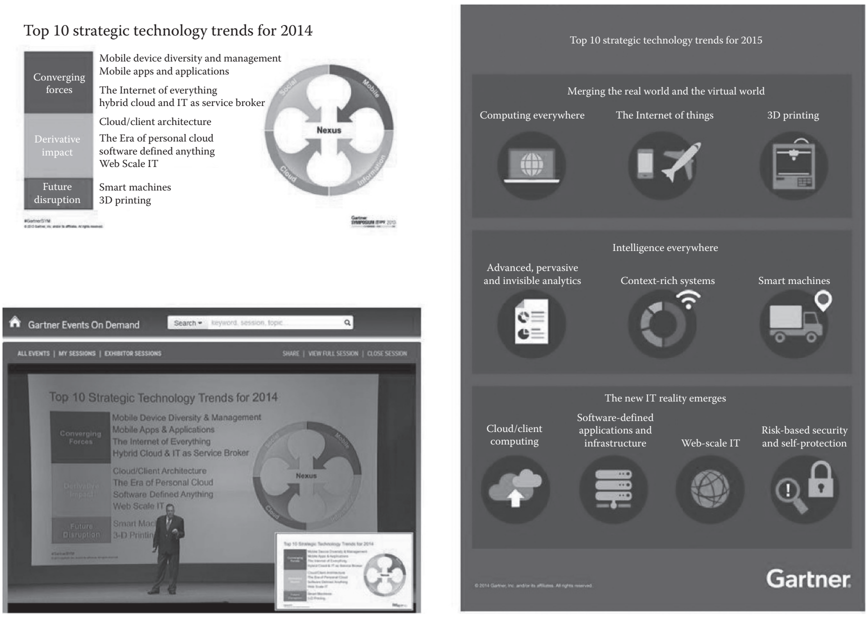

Figure 1.6 The top 10 strategic technologies in years 2014 and 2015.

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

In their prediction related to smart machines for 2014 and 2015, the following statements were made:

- By 2015, there will be more than 40 vendors with commercially available managed services offerings leveraging smart machines and industrialized services. By 2018, the total cost of ownership for business operations will be *reduced by 30%* through smart machines and industrialized services.
- Through 2020, the smart machine era will blossom with a proliferation of contextually aware, intelligent personal assistants, smart advisors (such as IBM Watson), advanced global industrial systems, and public availability of early examples of autonomous vehicles. The smart machine era will be the most disruptive in the history of IT.

In a recent report, *Cool Vendors in Smart Machines, 2014*, Gartner named three well-known examples of smart machines, including IBM's Watson, Google Now, and Apple's Siri. We will discuss few of the smart machines mentioned in the above predictions later in this chapter, but before that we will discuss *Deep Blue*, a chess-playing computer developed by IBM.

1.4.3 Deep Blue

In May 1997, IBM's Deep Blue became the first computer system to defeat the then-chess world champion Garry Kasparov in a match. The brute force of the computing power due to specialized hardware made Deep Blue capable of evaluating 200 million positions per second. The 259th most powerful supercomputer of 1997 was able to defeat the human world champion of chess. It was a historical achievement for the artificial intelligence community. How Deep Blue was able

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

to evaluate the situation on the chess board? The answer to this question is as follows [11]:

> Deep Blue's evaluation function was initially written in a generalized form, with many to-be-determined parameters (e.g. how important is a safe king position compared to a space advantage in the center, etc.). The optimal values for these parameters were then determined by the system itself, by analyzing thousands of master games. The evaluation function had been split into 8,000 parts, many of them designed for special positions. In the opening book there were over 4,000 positions and 700,000 grandmaster games. The endgame database contained many six piece endgames and five or fewer piece positions

In 1997, Deep Blue was a dedicated supercomputer against humans. Now the focus of research in the chess domain is to improve software efficiency, so that less powerful hardware is enough for the task. In 2006, chess program named *Deep Fritz* played a chess match against world champion *Vladimir Kramnik*. The program was executed on a personal computer containing two Intel Core 2 Duo CPUs. The program was capable of evaluating only 8 million positions per second as compared to the 200 million positions per second evaluation power of Deep Blue.

1.4.4 IBM's Watson

It was named after the first CEO of IBM, Thomas J. Watson. IBM's Watson is a wonderful machine that is capable of answering the questions posed in natural language.

Whether you call it a supercomputer, a cognitive computing system, or simply a question answering matching system—IBM Watson is perhaps the most well-known example of artificial intelligence in use today.

Watson gained its worldwide fame by receiving the first prize inquiz show “Jeopardy!”. With its supercomputing and AI power, Watson is able to help different industries by powering different types of practical applications. The industries benefiting from Watson include healthcare, finance, legal, and retail sector.

1.4.5 Google Now

Google’s innovation “Google Now” is another landmark for machine learning world. It is a personal assistant with an element of smartness and intelligence in it. The functions of Google Now include answering questions, making recommendations, and performing actions by assigning requests to a set of web services. With it, users can use voice commands to create reminders and get help with trivia questions. The proactive program observes the search habits of the users and uses them to predict the information that may be useful for users and delivers it to them.

1.4.6 Apple’s Siri

Siri (speech interpretation and recognition interface) is a widely used intelligent personal assistant by Apple Inc. Siri supports a number of languages including English, Spanish, French, German, Italian, Japanese, Korean, Mandarin, Russian, Turkish, and Arabic. Siri, just like any other personal assistant is updated to improve its response. Context understanding is very important. For example, if Siri is being told by a terrorist that he is going to blast a particular restuarant, Siri rather than showing the map of that restaurant, should respond by reporting such intention to some terrorism prevention center.

1.4.7 Microsoft’s Cortana

Microsoft’s Cortana is another intelligent personal assistant competing Google Now and Apple’s Siri. Users will be soon

International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

able to use Skype to book trips, shop, and plan their schedules, by chatting with Cortana.

1.5 Objective of This Book

The objectives of this book are as follows:

- Explanation of the concepts of machine learning algorithms
- Demonstration of simple practical example(s) to make the reader understand each algorithm

We believe that this book will be a very useful resource for beginners as well as researchers and IT security professional.

We have divided our books into two sections.

1. Supervised Learning Algorithms
2. Unsupervised Learning Algorithms

In the first section, we will discuss following algorithms:

1. Decision trees
2. Rule-based algorithms
3. Naïve Bayesian algorithm
4. Nearest neighbor algorithm
5. Neural networks
6. Linear discriminant analysis
7. Support vector machine

In the second section, we will discuss following algorithms:

1. *K*-means
2. Gaussian mixture model
3. Hidden Markov model
4. Principal components analysis in the context of dimensionality reduction

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)

References

1. Durant, W. Is man a machine? in *The Pleasures of Philosophy: A Survey of Human Life and Destiny*. New York: Simon and Schuster, 1953, p. 60.
2. Guthrie. Looking backwards, looking forwards: SAS, data mining and machine learning, 2014, http://blogs.sas.com/content/subconsciousmusings/2014/08/22/looking-backwards-looking-forwards-sas-data-mining-and-machine-learning/.
3. Mitchell, T. M. *The Discipline of Machine Learning*, Machine Learning Department technical report CMU-ML-06-108. Pittsburgh, PA: Carnegie Mellon University, July 2006.
4. Mitchell, T. M. *Machine Learning*. New York: McGraw-Hill, 1997.
5. Carroll, L. and Kelly, R. M. *Alice's Adventures in Wonderland*. Peterborough, ON: Broadview Press, 2000.
6. Bill Gates. AZQuotes.com, Wind and Fly LTD, 2016. http://www.azquotes.com/quote/850928 (accessed April 13, 2016).
7. United States Postal Service. https://about.usps.com/who-we-are/postal-facts/innovation-technologies.htm (accessed October 4, 2015).
8. Toyota Global Newsroom. http://http://newsroom.toyota.co.jp/en/detail/9753831 (accessed April 14, 2016).
9. Luhn, H. P. A business intelligence system, *IBM Journal* 2(4):314-319, 1958.
10. Vinyals, O. and Le, Q. V. A neural conversational model, *Proceedings of the 31st International Conference in Machine Learning*, Vol.37, arXiv:1506.05869v3, 2015.
11. Campbell, M., Hoane Jr. A. J., and Hsu, F.-H. Deep blue, *Artificial Intelligence* 134(1–2): 57-83, 2002.

Click here to order "Machine Learning: Algorithms and Applications"
International Standard Book Number-13: 978-1-4987-0538-7 (Hardback)