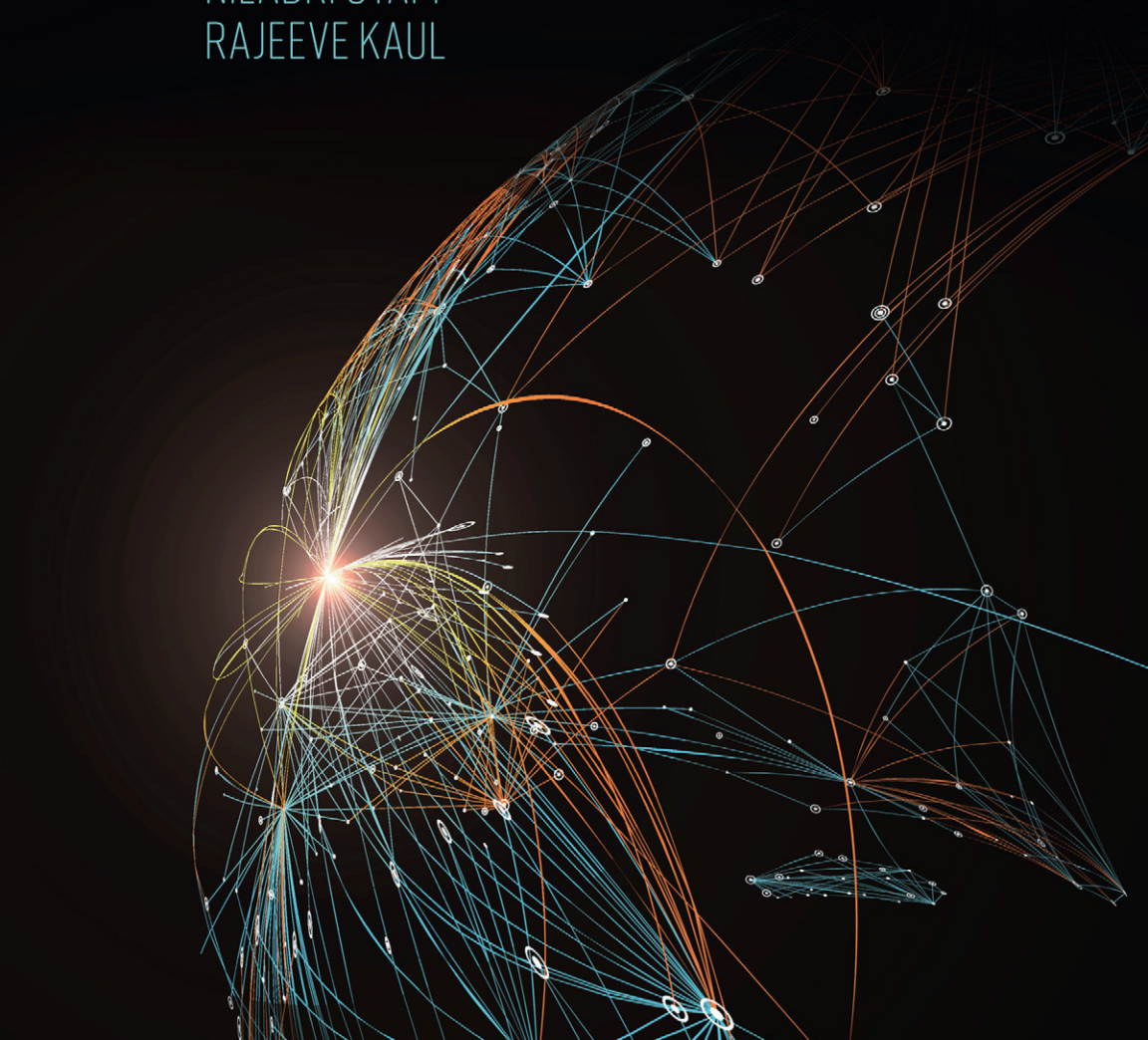
MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE IN MARKETING AND SALES

Essential Reference for Practitioners and Data Scientists

NILADRI SYAM
RAJEEVE KAUL



Machine Learning and Artificial Intelligence in Marketing and Sales

Praise for *Machine Learning and Artificial Intelligence in Marketing and Sales*

The world of business in general and marketing in particular is going through a radical transformation with the application of machine learning and artificial intelligence touching almost every aspect of business. These advanced tools are no longer solely the domain of statisticians and data scientists. Business practitioners find themselves interacting with these methods and, at the same time, data scientist often find themselves sitting at the management table managing groups and discussing their analysis. *Machine Learning and Artificial Intelligence in Marketing and Sales: Essential Reference for Practitioners and Data Scientists* strikes a, difficult to achieve, balance between providing sufficient information on commonly used but complex machine learning and AI tools, and yet keeping the book accessible and applicable to business practitioners with technical orientation. This is a great introduction book for those who wish to know not only about machine learning and AI, but also what it really is, and how to apply it in marketing and sales settings.

Oded Netzer, Professor of Business, Columbia University

This book is a great resource for Data scientists as a reference to anchor your technical understanding, build your intuition of the core machine learning models and at the same time elevate it for application in the real-world context of Marketing and Sales. It would be a good foundational book for students and applied practitioners of non-ML, non-stat backgrounds to gain confidence in your work and to stand behind the choices you make in your model building process. As the authors say, it does a good job at bridging the DS-real world application gap. I also liked the choice of the three models (NN, RF, SVM) to allow for focus and not overload the reader with tons of other material available. These three will get 80−90% of your job done as a modeler. I also liked the Executive Summary sections which appeal to the applied modeler in me and the Technical Detours which piqued a deeper intellectual curiosity and understanding of the details. Because they are positioned as detours, I could still read and use the book without making the technical parts overwhelming.

Vijay Jayanti, Head of Marketing Data Sciences at WhatsApp Inc.

In *Machine Learning and Artificial Intelligence in Marketing and Sales*, Syam and Kaul have teamed up to present a timely explanation of important topics in the evolving high-tech world of big data. For readers well-versed in the Support Vector

Machine, artificial neural nets, and deep learning, the book will be immediately useful. For readers new to these topics, the authors' accessible style lowers entry barriers. The book is required reading for managers, analysts, professors, and consultants involved in marketing and sales.
David J. Curry. Professor of Marketing, University of Cincinnati

Syam and Kaul's book is a comprehensive treatise on data science of marketing, a rich and deeply informative dive into the next generation of marketing analytics solutions. The work comprehensively integrates the theoretical concepts of Machine Learning with practical applications of marketing, making it essential for either ML Engineers solving marketing problems or marketing analysts looking to get a rigorous treatment of the nascent science.
Alex Vayner, Data science and AI expert, Partner, PA Consulting

The authors have skillfully tailored the content to a wide audience. I found this book as a solid reference guide for students and a reference for data science practitioners alike. While the book covers the most important Machine Learning topics in lucid detail, it also provides insightful executive summaries, and, most importantly, showcases applications of each model in the practical world of Sales and Marketing. I will wholeheartedly recommend this book to anyone interested in learning Machine Learning and Artificial Intelligence.
Sunish Mittal, Vice President, Data and Analytics, Aramark

This page intentionally left blank

Machine Learning and Artificial Intelligence in Marketing and Sales: Essential Reference for Practitioners and Data Scientists

BY

NILADRI SYAM
University of Missouri, USA

RAJEEVE KAUL
McDonald's Corporation, USA

United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2021

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-80043-881-1 (Print)
ISBN: 978-1-80043-880-4 (Online)
ISBN: 978-1-80043-882-8 (Epub)

ISOQAR certified Management System, awarded to Emerald for adherence to Environmental standard ISO 14001:2004.

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

This book is dedicated to the memory of Ma and Baba: NS

This book is dedicated to the memory of my parents who encouraged me to remember that learning is a lifelong process of staying abreast of change: RK

This page intentionally left blank

Table of Contents

List of Figures, Tables and Illustrations

Foreword

The book is written in five chapters covering important concepts, principles, and practices on contemporary machine learning topics. Each page is written in an easy-to-read format using clean and lean sentences unencumbered by complex jargon. Each chapter provides solid theoretical background of the methods selected, and further elaborates the how and the why aspects of model selection, model building and model validation; the step-by-step approaches of leveraging specific techniques such as the Neural Network, the decision trees, and the vector machines; and the pros and the cons of certain machine learning (ML) procedures. Moreover, each topic provides many real-world examples that connect the theory with the applied use. Lists and depth of supporting reference materials are also excellent.

We are at an exciting time where the era of big data, machine learning, artificial intelligence, cloud computing and advanced analytics is ushering unprecedented access to and uses of large volumes of data to improve our predictive power to unlock transformational changes that impact many aspects of our lives in the retail, the financial, the manufacturing, the technology, the healthcare, and other industries.

As both big and small firms alike are fine-tuning their pricing, promotion, distribution, customer-retention, risk-management, and go-to-market strategies, data scientists are increasingly expected to know cutting-edge solutions, equip themselves with many facets of ML techniques and solutions. This book undoubtedly provides the foundational background, the tools, and the necessary tips to grasp many of the ML methods currently in use. In addition, as ML is rapidly and dynamically evolving to impact our daily life, the timeliness of this book is undoubtedly very appropriate.

I have worked as a data scientist for diverse organizations and have taught analytics and ML classes in universities. A few pages into this book, I knew that it is a special treat for my appetite, and it really struck a chord. Many of the well-organized core concepts expounded in the book are not only refreshing but also the kind I wish I had long ago. I also dare to describe this book as a versatile tool and a must-have reference material for both beginners and seasoned data scientists alike, business leaders and those who embrace data and analytics-driven decision-making processes. In addition, analytics teachers and their respective students can benefit from the in-depth analysis of the contemporary data science topics and the plethora of examples provided. I commend the authors for a job well done.

Dawit Mulugeta, PhD (Biostatistics), VP Analytics and Risk Management, Wells Fargo

Dawit Mulugeta is an applied data scientist. Currently he holds a vice president of analytics and risk management role with Wells Fargo. In addition, he teaches popular analytics and ML classes in the Department of Management Sciences as well as in the Department of Accounting and Information Systems of the Max M. Fisher College of Business of the Ohio State University in Columbus. Dawit earned a PhD in biostatistics from the University of Wisconsin in Madison, USA.

Preface

Machine Learning and Artificial Intelligence in Marketing and Sales: Essential Reference for Practitioners and Data Scientists is intended for a variety of audiences:

- Marketing and sales practitioners who want to develop a deeper appreciation for how machine learning models can be applied to problems in marketing and sales.
- Data scientists who are tasked with implementing data-based solutions to business problems in the domain of marketing and sales.
- Software engineers and IT developers who will implement, or assist in the implementation of, and manage the solutions for marketing and sales organizations in their company.
- Students in master's programs in data science and in MBA programs and management consultants who wish to have a deeper appreciation of how machine learning and AI are impacting the world of marketing and sales.

This book represents a fruitful collaboration between an academic and an industry practitioner. Each author made a genuine attempt to understand and incorporate in the book the viewpoints and tastes of the other. This main purpose of this book is to bridge, what we call, the *Domain Specialist – Data Scientist Gap* (*DS-DS Gap*). It is aimed at the translators, that is, the boundary-spanners, among two distinct audiences – the marketing practitioners and the data scientists. It is the experience of the authors that often in companies one of the biggest barriers to success is the ability of the technical and sales/marketing business teams to effectively understand and communicate with one another in solving business problems. Marketing practitioners and data scientists wishing to work together to solve marketing and sales issues using the methods of machine learning need to have a shared understanding of how these methods can be applied to marketing and sales. This book treads the fine line between the very technical books on the one hand, and, on the other hand, the purely qualitative books that merely mention the various AI and machine learning applications in marketing and sales.

Any collaboration between an academic and an industry practitioner, with the former emphasizing theory and the latter emphasizing business applications, always has a fair bit of tension – the desirable kind of tension that has hopefully

made the final product better. We have tried earnestly to strike a balance between the theoretical/technical and applications aspects. Having said that, we have decided to err on the side of applications, anchoring our narrative on the connections between the techniques and their applications in a business setting. We have deliberately kept the technical details to a bare minimum in the main body of the text, and have dealt with technical details through the various "Technical detours" that have been collected together at the end of each chapter. This allows readers who do not wish to tackle the technical issues to be able to read the chapters easily without being distracted or overwhelmed by technical details. In addition, to help readers get a quick overview of the concepts involved, we have added "Executive summaries" as and when needed. As far as possible in each chapter we have tried to emphasize the intuitions behind the, sometimes complex, concepts of machine learning methods.

As far as the marketing and sales practitioners are concerned, the book assumes that they are interested in actual implementation of machine learning models, either by themselves or in collaboration with the data scientists in their organizations. For this reason, this book is not just a high-level overview of machine learning applications to marketing and sales. Thus, by its very nature the chapters assume that the reader is willing to handle some technical material. However, we have made every attempt to make the chapters self-contained by providing the background material needed to understand the chapter contents. Each chapter has a section on the existing applications of machine learning and artificial intelligence (AI) to issues in marketing and sales. We have tried to focus on applications that have been archived in the major peer-reviewed research journals in marketing, operations research, machine learning, expert systems, etc. By their nature, journal articles have details of implementation and data sets and interested readers can go through the articles listed in the references at the end of every chapter for further details. Finally, for those wishing to get hands-on experience of actually running analyses of marketing and sales data using machine learning models, each chapter has a couple of detailed "Case studies" at the end.

Acknowledgments

Niladri

I would like to gratefully acknowledge the support of my wife, Nivedita, without whose patience and encouragement this book would never have materialized. I would also like to thank my teacher, Professor Bibek Debroy, who introduced me to the power of quantitative modeling and testing of business phenomena in the 1980s, much before the term "analytics" had become fashionable. The *Center for Sales and Customer Development* (CSCD) at the University of Missouri provided the support and proper climate which greatly facilitated the writing of this book.

Rajeeve

This book would not be possible without the unyielding support and encouragement of my wife, Shalini, and the patience and understanding of our son, Harsha. Working on a book while staying abreast with challenging executive roles required them to sacrifice time that we could have spent building life memories – for which I am so grateful. I would also like to thank my many professors who encouraged me to learn and experiment with so many quantitative methods across diverse fields from statistics, to marketing, finance, operations research, etc. I further extend my gratitude to the many incredible executives across so many industries who adopted my quantitative solutions to improve decision in areas including pricing, marketing, supply chain,and digital among others, and the companies that allowed me to follow my curiosity to develop and deploy these models.

This page intentionally left blank

Introduction

This book grew out of many discussions that we, the two coauthors, had over the span of several years. Over the years, the field of machine learning has gone from an esoteric topic discussed among a small select group of practitioners and researchers to an ever-growing tsunami of interest and practitioners approaching an increasingly digitized world from their diverse backgrounds. We were both interested in machine learning, but we had approached it from two very different perspectives, and could relate to this heterogeneity in thinking. One of us is an academic and was focused primarily on the theory of machine learning and in doing research on this topic. The other author is an industry practitioner and was focused primarily on the applications of machine learning models in marketing and sales. Of course, despite the different areas of emphasis each one of us is interested in both theories and applications, and both realize that these should go hand in hand. As our discussions progressed, we felt that the existing resources in machine learning did not quite serve the needs of the diverse stakeholders that have to work together for successful industry applications of machine learning in marketing and sales. This motivated the need for a book that would speak to both the sales/marketing business teams and the data scientists who are tasked with solving business problems in the domains of marketing and sales.

This book takes a different approach compared to two distinct existing categories of books that deal with machine learning and AI – the technical and the qualitative books. The former category of books does not focus on applications in a specific domain in any detail, and often use stylized examples drawn not from business but from the physical sciences. The latter (qualitative) category of books does not provide any details of the statistical and mathematical concepts that drive machine learning techniques and their applications. Neither of these types of books serve well the needs of practitioners coming from varied backgrounds working on actual implementation of machine learning models in the field of sales and marketing in a business enterprise. As far as the technical aspects are concerned, we have avoided machine learning algorithms and have focused instead on the concepts and ideas that underlie machine learning models and methods. Our interest is in connecting the concepts that underpin these methods and bridge the gap between the data scientist and the business

Machine Learning and Artificial Intelligence in Marketing and Sales, 1–3

doi:10.1108/978-1-80043-880-420211001

practitioner. There are many online and other resources for those readers wishing to familiarize themselves with algorithms.

A major decision in writing any book is always what to include and what to leave out. Machine learning and AI are flourishing fields of research with scholars from diverse disciplines such as applied mathematics, statistics, operations research, engineering, and computer science actively contributing to them. There is an enormous variety of machine learning models, and trying to include all these models would make the book unwieldy and unfit for the non-expert. We have therefore decided to focus on just three of the most commonly used methods in marketing and sales applications – Neural Network, Support Vector Machine (SVM), and Random Forest. A key motivation for this approach is the acknowledgment that though there are many ideas and approaches, not all of them have efficacy across a broad set of business problems. As such, it makes sense to focus on methods that have been validated for their applicability in solving a wide variety of marketing problems. Importantly, these three models are exemplars of three distinct classes of machine learning models, and many of the latest developments in the field are based on these three fundamental models. Thus, any understanding of the latest developments in machine learning, deep learning, and AI require an understanding of these models. For example, advances in deep learning including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are based on the fundamental ideas of Neural Networks. An SVM is a prominent example of the class of kernel-based machine learning models. A Random Forest is a good representative of the class of tree-based learning models and is a good context to discuss ideas of bagging, boosting and gradient boosting.

This book is about machine learning and Artificial Intelligence (AI). While different authors have different ideas about the distinction between them, the consensus opinion is that machine learning is a subfield of AI. At a broad level, AI is the umbrella term used to denote the entire suite of technologies that are designed to mimic human abilities. Thus, AI includes machine learning, deep learning, natural language processing (NLP) etc. Many authors argue that neural networks are part of deep learning since the latter are essentially 'big' neural networks with many layers with complex interconnections between them. To the extent that machine learning and deep learning are often identified as distinct subfields of AI (see, for instance, the *SAS Institute* white paper by Thompson, Li, and Bolen), the question of whether neural networks should be included under machine learning often becomes a matter of taste and preferences. We would like to avoid these matters of semantics, and hence we have included both machine learning and AI in the title. The reader can think of the content of the book as "narrow AI" which is supervised machine learning.

Before discussing the three specific machine learning models, we will discuss the concepts of training and performance assessment since these concepts are applicable to all machine learning models. This is done in Chapter 1. Here we also discuss the linear regression model for a continuous dependent variable (often called a response variable or a target variable in a machine learning context) and

the logistic regression model for a categorical dependent variable. These will form useful benchmarks for the machine learning models discussed in Chapter 2 (Neural Networks), Chapter 4 (Support Vector Machine) and Chapter 5 (Random Forest). In Chapter 3 we discuss the very important concept of overfitting and regularization. We have decided to introduce these after the chapter on Neural Networks since it is easier to grasp these concepts when discussed in the context of a specific model, even though they are applicable for all models.

This page intentionally left blank

Chapter 1

Introduction and Machine Learning Preliminaries: Training and Performance Assessment

Chapter Outline

1. Training of Machine Learning Models

In this chapter, we will restrict our discussion to models that have a specific *response* variable. Response variables are also called *target* variables and machine learning models with such variables are known as *supervised learning*. These models are distinguished from unsupervised learning models, like clustering models, which do not have pre-specified response variables. We first describe briefly two categories of supervised learning models that are of interest to us – regression models and classification models. They are distinguished by the "type" of response variable.

Machine Learning and Artificial Intelligence in Marketing and Sales, 5–24

doi:10.1108/978-1-80043-880-420211002

1.1 Regression and Classification Models

Regression and classifications models are discussed in almost all statistics textbooks and we will not repeat these details here. We only mention them very briefly to set the stage for the discussion of machine learning models in the later chapters of our book.

Regression models have a continuous response variable (often called a *dependent* variable). We will consider the case of a *linear regression*. Consider the case of a consumer products company that provides free samples to consumers to induce trial and also word-of-mouth to sell its products. Sometimes these companies may have their salespeople stationed at various retailers to distribute their samples in the hope that, after trying it, the consumers will like the product and purchase the product after their initial trial ("trial-and-repeat" purchase models in marketing). There is obviously some time lag between trial and repeat, and suppose the company wants to understand how their distribution of samples in a given month induces repeat purchases in the next month. We will denote the number of samples in a given month, say, October, as X and the number of repeat purchases in November by Y. We can treat the number of purchases Y as a continuous variable. Thus, Y is the continuous response variable and X is the *explanatory* (also called *independent*) variable. A simple model to predict Y based on X is

$$Y = w_0 + w_1 X + \varepsilon \tag{1.1}$$

The epsilon term (ε) at the end is the error term. It captures the fact that the relationship between X and Y has randomness owing to a host of factors. The common sources of randomness are the many other factors that also affect purchases in November apart from trials in October. Of course, these have not been modeled, and thus, there will be errors when we use only one explanatory variable to predict purchases in November. In the simple linear regression above, the effect of the number of trial samples in October is given by the *parameter* w_1 (parameters that multiply inputs are also called *coefficients* and in machine learning models like Neural Networks, they are called *weights*). The slope, given by w_1, intuitively captures the additional purchases in November due to an extra trial sample in October. The intercept, given by w_0, intuitively captures the purchases in November if there were no trial samples in October (in machine learning models like Neural Networks this parameter is called the *bias*). Instead of just one explanatory variable, one could include other variables as well on the right-hand side of the equation, and then we would have a multiple regression.

In this book, we will refer to a model with a continuous response variable as a *regression model* and different machine learning techniques can be used to analyze such models. The traditional linear regression described above can serve as a useful benchmark to compare with the more recent machine learning models.

In marketing the response variable we are often interested in is categorical. For instance, consider the case of a bank that wants to predict whether its customers are likely to churn (leave) or not. A sales organization may be interested in

categorizing their prospects as being either in the "buy" or "not buy" category. In *lead scoring*, a sales organization may want to categorize their sales leads as belonging to one of many different classes based on their propensities to buy: very unlikely, unlikely, likely, very likely. These are classification tasks, with the first two being binary classification and the third being multiclass classification.

We will briefly describe the case of binary classification. The traditional workhorse for analyzing models with a binary categorical response variable is a *logistic regression*. In the bank churn example, suppose the two classes are "churn" or "not churn," and the bank wants to understand to what extent the amount of "balance" that the customer has is predictive of churn. The answer is not clear a priori. On the one hand, a customer with a large balance can be considered as having a deeper relationship with the bank, and therefore, less likely to churn. On the other hand, such attractive customers are targets of competitive offers from other banks and are more likely to churn. We use the balance a customer has in the bank as the explanatory variable X. The response variable $Y = \{+1, -1\}$ is coded as: $+1 \equiv$ "churn" and $-1 \equiv$ "not churn." We cannot use a linear regression here since we would like to model the probability of churning, and unlike the continuous response of a linear regression which can take on any value, probabilities have to lie in the interval [0, 1].

The logistic regression works by defining $p = \text{Probability}(Y = +1)$, and then positing the relationship

$$\text{Log}[p/(1-p)] = w_0 + w_1 X_1 \tag{1.2}$$

The term on the left-hand side, $\text{Log}[p/(1-p)]$, is called the log *odds ratio*. This formulation generates the probability of churning, p. It also ensures that the sum, Probability("churn") + Probability ("not churn"), adds up to 1 as is expected of probabilities. Based on these probabilities, one can classify customers as belonging to the category "churn" ("not churn") if $p > 0.5$ ($p < 0.5$).

In this book, we will refer to a model with a categorical response variable, both binary and multiclass, as a *classification model*, and various machine learning models can be used for classification tasks. The logistic regression described above can serve as a benchmark to compare with machine learning classification models.

1.2 Cost Functions and Training of Machine Learning Models

Machine learning practitioners often talk of *cost functions*. Take the example of a company trying to predict sales of a certain product. Data are available over many past periods, and in each period, sales are affected by factors like the company's own price, advertising spending, and the competitor's prices among other factors. Given this situation, we want to accurately predict the sales of the company. One way to do this is to create a mathematical formulation (model) that allows us to predict sales based on observed factors (like price, advertising, etc.) for each recorded period in the past. Then, we can compare the *actual* past sales value against the sales value *predicted* by this model to see how well the

model is performing. In this case, the cost function is a function of the difference between the predicted output of the model and the actual sales value for all past periods. The model is said to perform well when the cost (also called *error* or *loss*) is minimized. The minimization of cost is achieved by choosing appropriate parameters of the mathematical model. This process is called *training* the machine learning model.

For a machine learning model, training is said to occur when the model estimates the "best" values of the parameters. What does best mean? At this point, we formalize the concept of a cost function a bit more. Consider the linear regression model specified above. Given a specific input data point, $X = x$, and some values of the parameters (weights), the regression model can make a prediction $f(x)$.[1] This is, given specific values of w_0 and w_1 and a data point x, the regression model makes a prediction $y = f(x) = w_0 + w_1 x$. On the other hand, the input data point x has an actual observed y (also called *target*) associated with it. Intuitively, the cost function measures the discrepancy between the model prediction y and the actual y for all possible values of input x. The goal of training is to choose those parameters (weights w_0 and w_1) that minimize this cost. These cost minimizing weights are the "best" weights.

In our discussions earlier, the cost function was based on sales – specifically it was the difference between actual observed sales and the sales predicted by the model. In business, typical "performance indicators" one encounters are sales, margins, inventory balances, profits, hours worked, and payroll to name a few. Any of these, or any combination of these, could be used to define the cost function.

By far the most common technique for training most machine learning models is to use the method of *Maximum Likelihood Estimation* (MLE). Maximum likelihood estimators have desirable statistical properties and are therefore advantageous to use. Elements of this philosophy are also applied extensive in deep learning. It is noteworthy that there is a close theoretical connection between cost (loss) functions and maximum likelihood, and therefore we will use the maximum likelihood framework to address the issue of choosing appropriate cost functions. A standard result from statistics is that, when the errors in a regression model are Gaussian then minimizing the sum-of-squares cost with respect to the weights is equivalent to maximizing the log-likelihood. In the maximum likelihood framework, the appropriate cost function for regression-type outputs is the *sum of squares* cost (loss), and for binary output the appropriate cost is the *cross-entropy* cost. Expressions for the sum-of-squares cost function for regression and the cross-entropy cost function for binary classification are in the appendix.

Technical Detour 1

[1]The uppercase denotes the variable and the lowercase is a specific value of that variable.