Machine Learning And Causal Inference A Modular Approach

Random forests are a powerful method for non-parametric regression, but are limited in their ability to fit smooth signals. Taking the perspective of random forests as an adaptive kernel method, we pair the forest kernel with a local linear regression adjustment to better capture smoothness. The resulting procedure, local linear forests, enables us to improve on asymptotic rates of convergence for random forests with smooth signals, and provides substantial gains in accuracy on both real and simulated data. We prove a central limit theorem valid under regularity conditions on the forest and smoothness constraints, propose a computationally efficient construction for confidence intervals, and discuss an extension to local linear causal forests for learning heterogeneous treatment effects. Following this deep dive into local linear forests, we discuss two applications of machine learning for causal inference. The first is a retirement reform in Denmark, in which shifting eligibility ages for an early retirement program provide an opportunity to analyze heterogeneous treatment effects of the age of retirement eligibility. The second is a randomized controlled trial in Nairobi, Kenya, aiming to lower rates of gender-based violence against adolescent students living in informal settlements. In the latter example, we explore how local linear causal forests help to uncover and emphasize trends in marginalized student responses to the intervention. In both cases, we address how machine learning and causal inference are powerful tools to discover patterns in individual treatment effects, and to advocate for marginalized groups when estimates reveal troubling patterns in the data.

As healthcare data becomes increasingly ubiquitous, improving data-driven biomedical research is timely and important. There is a rush to learn from these new sources of data, and to implement research findings into clinical practice. While machine learning methods provide compelling examples of recognizing sophisticated patterns in data, their impact rests heavily on their ability to use data to influence decision making, especially in healthcare. The relationship between machine learning and decision making becomes particularly clear through the lens of causal inference. In general, the harm and benefit attributed to a medical decision depends on the causal treatment effect of the decision in the appropriate population, beyond their baseline risk of poor outcomes. In precision medicine research, the goal is to develop treatment decisions for individual patients by considering the sub-population of individuals with similar covariates to each patient. This thesis advances methodology and practice for applying machine learning to learn better decision-making rules that influence clinical practice, and understanding the fundamental possibilities and limitations of using data to learn to make optimal decisions. First, we develop an approach for personalized treatment effect estimation based on the relative ratio of treatment outcomes. Second, we study when we can trust causal results learned from data, and develop a sensitivity analysis for conditional and average treatment effects to bound the bias created from unobserved confounding. Third, noting that treatment benefit is highly correlated with baseline risk for preventative treatments for atherosclerotic cardiovascular disease (ASCVD), we use machine learning approaches to improve ASCVD risk predictions from longitudinal cohort data that affect clinical prescribing practice, particularly among under-represented minorities.

David A. Freedman presents a definitive synthesis of his approach to statistical modeling and causal inference in the social sciences.

Elements of Causal InferenceFoundations and Learning AlgorithmsMIT Press

Molecular descriptors are commonly used to digitally represent the physical structure of a molecule in quantitative structure-activity relationship machine learning models, but the overwhelming abundance of features can negatively impact the model's predictive performance. This research explores causal inference modeling as a method of feature engineering for ensemble-based machine learning techniques, and evaluates model performance against other common methods of feature evaluation.

This dissertation consists of three chapters that study causal inference when applying machinelearning methods. In Chapter 1, I propose an orthogonal extension of the semiparametric difference-in-differences estimator proposed in Abadie (2005). The proposed estimator enjoys the so-called Neyman-orthogonality (Chernozhukov et al. 2018) and thus it allows researchers to flexibly use a rich set of machine learning (ML) methods in the first-step estimation. It is particularly useful when researchers confront a high-dimensional data set when the number of potential control variables is larger than the sample size and the conventional nonparametric estimation methods, such as kernel and sieve estimators, do not apply. I apply this orthogonal difference-in-differences estimator to evaluate the effect of tariff reduction on corruption. The empirical results show that tariff reduction decreases corruption in large magnitude. In Chapter 2, I study the estimation and inference of the mode treatment effect. Mean,median, and mode are three essential measures of the centrality of probability distributions. In program evaluation, the average treatment effect (mean) and the quantile treatment effect (median) have been intensively studied in the past decades. The mode treatment effect, however, has long been neglected in program evaluation. This paper fills the gap by discussing both the estimation and inference of the mode treatment effect. I propose both traditional kernel and machine learning methods to estimate the mode treatment effect. I also derive the asymptotic properties of the proposed estimators and find that both estimators follow the asymptotic normality but with the rate of convergence slower than the regular rate N^1/2, which is different from the rates of the classical average and quantile treatment effect estimators. In Chapter 3 (joint with Liqiang Shi), we study the estimation and inference of the doublyrobust extension of the semiparametric quantile treatment effect estimation discussed in Firpo (2007). This proposed estimator allows researchers to use a rich set of machine learning methods in the first-step estimation, while still obtaining valid inferences. Researchers can include as many control variables as they consider necessary, without worrying about the over-fitting problem which frequently happens in the traditional estimation methods. This paper complements Belloni et al. (2017), which provided a very general framework to discuss the

estimation and inference of many different treatment effects when researchers apply machine learning methods.
A concise and self-contained introduction to causal inference, increasingly important in data science and machine learning. The mathematization of causality is a relatively recent development, and has become increasingly important in data science and machine learning. This book offers a self-contained and concise introduction to causal models and how to learn them from data. After explaining the need for causal models and discussing some of the principles underlying causal inference, the book teaches readers how to use causal models: how to compute intervention distributions, how to infer causal models from observational and interventional data, and how causal ideas could be exploited for classical machine learning problems. All of these topics are discussed first in terms of two variables and then in the more general multivariate case. The bivariate case turns out to be a particularly hard problem for causal learning because there are no conditional independences as used by classical methods for solving multivariate cases. The authors consider analyzing statistical asymmetries between cause and effect to be highly instructive, and they report on their decade of intensive research into this problem. The book is accessible to readers with a background in machine learning or statistics, and can be used in graduate courses or as a reference for researchers. The text includes code snippets that can be copied and pasted, exercises, and an appendix with a summary of the most important technical concepts.
Machine learning tools are well known for their success in prediction. But prediction is not causation, and causal discovery is at the core of most questions concerning economic policy. Recently, however, the literature has focused more on issues of causality. This paper gently introduces some leading work in this area, using a concrete example—assessing the impact of a hypothetical banking crisis on a country’s growth. By enabling consideration of a rich set of potential nonlinearities, and by allowing individually-tailored policy assessments, machine learning can provide an invaluable complement to the skill set of economists within the Fund and beyond.
This textbook for graduate students in statistics, data science, and public health deals with the practical challenges that come with big, complex, and dynamic data. It presents a scientific roadmap to translate real-world data science applications into formal statistical estimation problems by using the general template of targeted maximum likelihood estimators. These targeted machine learning algorithms estimate quantities of interest while still providing valid inference. Targeted learning methods within data science area critical component for solving scientific problems in the modern age. The techniques can answer complex questions including optimal rules for assigning treatment based on longitudinal data with time-dependent confounding, as well as other estimands in dependent data structures, such as networks. Included in Targeted Learning in Data Science are demonstrations with soft ware packages and real data sets that present a case that targeted learning is crucial for the next generation of statisticians and data scientists. Th is book is a sequel to the first textbook on machine learning for causal inference, Targeted Learning, published in 2011. Mark van der Laan, PhD, is Jiann-Ping Hsu/Karl E. Peace Professor of Biostatistics and Statistics at UC Berkeley. His research interests include statistical methods in genomics, survival analysis, censored data, machine learning, semiparametric models, causal inference, and targeted learning. Dr. van der Laan received the 2004 Mortimer Spiegelman Award, the 2005 Van Dantzig Award, the 2005 COPSS Snedecor Award, the 2005 COPSS Presidential Award, and has graduated over 40 PhD students in biostatistics and statistics. Sherri Rose, PhD, is Associate Professor of Health Care Policy (Biostatistics) at Harvard Medical School. Her work is centered on developing and integrating innovative statistical approaches to advance human health. Dr. Rose’s methodological research focuses on nonparametric machine learning for causal inference and prediction. She co-leads the Health Policy Data Science Lab and currently serves as an associate editor for the Journal of the American Statistical Association and Biostatistics.
Das maschinelle Lernen ist zwangsläufi g eines der am schnellsten wachsenden Gebiete der Computerwissenschaft. Nicht nur die zu verarbeitenden Datenmengen werden immer umfangreicher, sondern auch die Theorie, wie man sie verarbeiten und in Wissen verwandeln kann. Maschinelles Lernen ist ein verständlich geschriebenes Lehrbuch, welches ein breites Spektrum an Themen aus verschiedenen Bereichen abdeckt, wie zum Beispiel Statistik, Mustererkennung, neuronale Netze, künstliche Intelligenz, Signalverarbeitung, Steuerung und Data Mining. Darüber hinaus beinhaltet das Buch auch Themen, die von einführenden Werken häufi g nicht behandelt werden. Unter anderem: Überwachtes Lernen; Bayessche Entscheidungstheorie; parametrische und nichtparametrische Statistik; multivariate Analysis; Hidden-Markow-Modelle; bestärkendes Lernen; Kernel-Maschinen; graphische Modelle; Bayes-Schätzung und statistischen Testmethoden. Da maschinelles Lernen eine immer größere Rolle für Studierende der Informatik spielt, geht die zweite Aufl age des Buches auf diese Veränderung ein und unterstützt gezielt Anfänger in diesem Gebiet, unter anderem durch Übungsaufgaben und zusätzlichen Beispieldatensätzen. Prof. Dr. Ethem Alpaydin, Bogaziçi University, Istanbul.
Alles hat sich geändert, als der Zeiger des Weltalters von 19 auf 20 sprang. Auf fast allen Gebieten wurden im 20. Jahrhundert Entdeckungen gemacht oder Ideen entwickelt, die unser Bild vom Universum und von uns selbst auf den Kopf gestellt haben. Alles schien neu, nichts unmöglich: Maschinen, die denken, Hunde im Weltall und Menschen auf dem Mond. Alte Gewissheiten büßten ihre Geltung ein, hergebrachte Autoritäten verloren ihre Macht. Die Welt wollte kein Zentrum mehr kennen. Auf ganz eigene Weise führt John Higgs durch dieses Jahrhundert der Genies und Gurus. Er erläutert die Relativitätstheorie anhand eines fallenden Würstchens, erzählt von Satanisten im Raumfahrtprogramm der Amerikaner und geht der Frage nach, ob ein Schmetterling in Brasilien einen Tornado in Texas auslösen kann. Das ist alles unglaublich seltsam und ziemlich wahnsinnig. Ein Buch wie ein Trip.
Statistical Methods for Dynamic Treatment Regimes shares state of the art of statistical methods developed to address questions of estimation and inference for dynamic treatment regimes, a branch of personalized medicine. This volume demonstrates these methods with their conceptual underpinnings and illustration through analysis of real and simulated data. These methods are immediately applicable to the practice of personalized medicine, which is a medical paradigm that emphasizes the systematic use of individual patient information to optimize patient health care. This is the first single

source to provide an overview of methodology and results gathered from journals, proceedings, and technical reports with the goal of orienting researchers to the field. The first chapter establishes context for the statistical reader in the landscape of personalized medicine. Readers need only have familiarity with elementary calculus, linear algebra, and basic large-sample theory to use this text. Throughout the text, authors direct readers to available code or packages in different statistical languages to facilitate implementation. In cases where code does not already exist, the authors provide analytic approaches in sufficient detail that any researcher with knowledge of statistical programming could implement the methods from scratch. This will be an important volume for a wide range of researchers, including statisticians, epidemiologists, medical researchers, and machine learning researchers interested in medical applications. Advanced graduate students in statistics and biostatistics will also find material in Statistical Methods for Dynamic Treatment Regimes to be a critical part of their studies.

In this thesis, I address three challenging machine-learning problems. The first problem that we address is the imbalanced data problem. We propose two algorithms to handle highly imbalanced classification problems. The first algorithm uses mixed integer programming to optimize a weighted balance between positive and negative class accuracies. The second method uses an approximation in order to assist with scalability. Specifically, it follows a characterize-then-discriminate approach. The positive class is first characterized by boxes, and then each box boundary becomes a separate discriminative classifier. This method is computationally advantageous because it can be easily parallelized, and considers only the relevant regions of the feature space. The second problem is a density estimation problem for categorical data sets. We present tree- and list- structured density estimation methods for binary/categorical data. We present three generative models, where the first one allows the user to specify the number of desired leaves in the tree within a Bayesian prior. The second model allows the user to specify the desired number of branches within the prior. The third model returns lists (rather than trees) and allows the user to specify the desired number of rules and the length of rules within the prior. Finally, we present a new machine learning approach to estimate personalized treatment effects in the classical potential outcomes framework with binary outcomes. Strictly, both treatment and control outcomes must be measured for each unit in order to perform supervised learning. However, in practice, only one outcome can be observed per unit. To overcome the problem that both treatment and control outcomes for the same unit are required for supervised learning, we propose surrogate loss functions that incorporate both treatment and control data. The new surrogates yield tighter bounds than the sum of the losses for the treatment and control groups. A specific choice of loss function, namely a type of hinge loss, yields a minimax support vector machine formulation. The resulting optimization problem requires the solution to only a single convex optimization problem, incorporating both treatment and control units, and it enables the kernel trick to be used to handle nonlinear (also non-parametric) estimation.

This thesis discusses methods for drawing causal inference in the presence of data pathologies such as missingness and multiple hypotheses, motivated by statistical problems encountered in everyday public health and medical practice. The thesis consists of three parts, addressing problems arising from malaria, human immunodeficiency virus (HIV) and trauma data.

Artificial Intelligence and Causal Inference address the recent development of relationships between artificial intelligence (AI) and causal inference. Despite significant progress in AI, a great challenge in AI development we are still facing is to understand mechanism underlying intelligence, including reasoning, planning and imagination. Understanding, transfer and generalization are major principles that give rise intelligence. One of a key component for understanding is causal inference. Causal inference includes intervention, domain shift learning, temporal structure and counterfactual thinking as major concepts to understand causation and reasoning. Unfortunately, these essential components of the causality are often overlooked by machine learning, which leads to some failure of the deep learning. AI and causal inference involve (1) using AI techniques as major tools for causal analysis and (2) applying the causal concepts and causal analysis methods to solving AI problems. The purpose of this book is to fill the gap between the AI and modern causal analysis for further facilitating the AI revolution. This book is ideal for graduate students and researchers in AI, data science, causal inference, statistics, genomics, bioinformatics and precision medicine. Key Features: Cover three types of neural networks, formulate deep learning as an optimal control problem and use Pontryagin's Maximum Principle for network training. Deep learning for nonlinear mediation and instrumental variable causal analysis. Construction of causal networks is formulated as a continuous optimization problem. Transformer and attention are used to encode-decode graphics. RL is used to infer large causal networks. Use VAE, GAN, neural differential equations, recurrent neural network (RNN) and RL to estimate counterfactual outcomes. AI-based methods for estimation of individualized treatment effect in the presence of network interference.

The use of Electronic Health Records (EHR)/Electronic Medical Records (EMR) data is becoming more prevalent for research. However, analysis of this type of data has many unique complications due to how they are collected, processed and types of questions that can be answered. This book covers many important topics related to using EHR/EMR data for research including data extraction, cleaning, processing, analysis, inference, and predictions based on many years of practical experience of the authors. The book carefully evaluates and compares the standard statistical models and approaches with those of machine learning and deep learning methods and reports the unbiased comparison results for these methods in predicting clinical outcomes based on the EHR data. Key Features: Written based on hands-on experience of contributors from multidisciplinary EHR research projects, which include methods and approaches from statistics, computing, informatics, data science and clinical/epidemiological domains. Documents the detailed experience on EHR data extraction, cleaning and preparation Provides a broad view of statistical approaches and machine learning prediction models to deal with the challenges and limitations of EHR data. Considers the complete cycle of EHR data analysis. The use of EHR/EMR analysis requires close collaborations between statisticians, informaticians, data

scientists and clinical/epidemiological investigators. This book reflects that multidisciplinary perspective.
This book compiles leading research on the development of explainable and interpretable machine learning methods in the context of computer vision and machine learning. Research progress in computer vision and pattern recognition has led to a variety of modeling techniques with almost human-like performance. Although these models have obtained astounding results, they are limited in their explainability and interpretability: what is the rationale behind the decision made? what in the model structure explains its functioning? Hence, while good performance is a critical required characteristic for learning machines, explainability and interpretability capabilities are needed to take learning machines to the next step to include them in decision support systems involving human supervision. This book, written by leading international researchers, addresses key topics of explainability and interpretability, including the following: · Evaluation and Generalization in Interpretable Machine Learning · Explanation Methods in Deep Learning · Learning Functional Causal Models with Generative Neural Networks · Learning Interpreatable Rules for Multi-Label Classification · Structuring Neural Networks for More Explainable Predictions · Generating Post Hoc Rationales of Deep Visual Classification Decisions · Ensembling Visual Explanations · Explainable Deep Driving by Visualizing Causal Attention · Interdisciplinary Perspective on Algorithmic Job Candidate Search · Multimodal Personality Trait Analysis for Explainable Modeling of Job Interview Decisions · Inherent Explainability Pattern Theory-based Video Event Interpretations
Exploiting and learning graph structures is becoming ubiquitous in Network Information Theory and Machine Learning. The former deals with efficient communication schemes in a many-node network. In the latter, inferring graph structured relationships from high dimensional data is important. In this dissertation, some graph theoretic results in these two areas are presented. The first part deals with the problem of optimizing bandwidth resources for a shared broadcast link serving many users each having access to cached content. This problem and its variations are broadly called Index Coding. Index Coding is fundamental to understanding multi-terminal network problems and has applications in networks that deploy caches. The second part deals with the resources required for learning a network structure that encodes distributional and causal relationships among many variables in machine learning. The number of samples needed to learn graphical models that capture crucial distributional information is studied. For learning causal relationships, when passive data acquisition is not sufficient, the number of interventions required is investigated. In the first part, efficient algorithms for placing popular content in a network that deploys a distributed system of caches are provided. Then, the Index Coding problem is considered: every user has its own cache content that is given and transmissions on a shared link are to be optimized. All graph theoretic schemes for Index Coding, known prior to this work, are shown to perform within a constant factor from the one based on graph coloring. Then, `partial' flow-cut gap results for information flow in a multi-terminal network are obtained by leveraging Index Coding ideas. This provides a poly-logarithmic approximation for a known generalization of multi-cut. Finally, optimal cache design in Index Coding for an adversarial demand pattern is considered. Near-optimal algorithms for cache design and delivery within a broad class of schemes are presented. In the second part, sample complexity lower bounds considering average error for learning random Ising Graphical Models, sampled from Erdós-Rényi ensembles, are obtained. Then, the number of bounded interventions required to learn a network of causal relationships under the Pearls model is studied. Upper and lower bounds on the number of size bounded interventions required for various classes of graphs are obtained.
This volume seeks to infer large phylogenetic networks from phonetically encoded lexical data and contribute in this way to the historical study of language varieties. The technical step that enables progress in this case is the use of causal inference algorithms. Sample sets of words from language varieties are preprocessed into automatically inferred cognate sets, and then modeled as information-theoretic variables based on an intuitive measure of cognate overlap. Causal inference is then applied to these variables in order to determine the existence and direction of influence among the varieties. The directed arcs in the resulting graph structures can be interpreted as reflecting the existence and directionality of lexical flow, a unified model which subsumes inheritance and borrowing as the two main ways of transmission that shape the basic lexicon of languages. A flow-based separation criterion and domain-specific directionality detection criteria are developed to make existing causal inference algorithms more robust against imperfect cognacy data, giving rise to two new algorithms. The Phylogenetic Lexical Flow Inference (PLFI) algorithm requires lexical features of proto-languages to be reconstructed in advance, but yields fully general phylogenetic networks, whereas the more complex Contact Lexical Flow Inference (CLFI) algorithm treats proto-languages as hidden common causes, and only returns hypotheses of historical contact situations between attested languages. The algorithms are evaluated both against a large lexical database of Northern Eurasia spanning many language families, and against simulated data generated by a new model of language contact that builds on the opening and closing of directional contact channels as primary evolutionary events. The algorithms are found to infer the existence of contacts very reliably, whereas the inference of directionality remains difficult. This currently limits the new algorithms to a role as exploratory tools for quickly detecting salient patterns in large lexical datasets, but it should soon be possible for the framework to be enhanced e.g. by confidence values for each directionality decision.
This state-of-the-art survey is dedicated to the memory of Emmanuil Markovich Braverman (1931-1977), a pioneer in developing machine learning theory. The 12 revised full papers and 4 short papers included in this volume were presented at the conference "Braverman Readings in Machine Learning: Key Ideas from Inception to Current State" held in Boston, MA, USA, in April 2017, commemorating the 40th anniversary of Emmanuil Braverman's decease. The papers present an overview of some of Braverman's ideas and approaches. The collection is divided in three parts. The first part bridges the past and the present and covers the concept of kernel function and its application to signal and image analysis as well as clustering. The second part presents a set of extensions of Braverman's work to issues of current interest both in theory and applications of machine learning. The third part includes short essays by a friend, a student, and a colleague.
This dissertation focuses on the prediction of agricultural land values and the effects of water rights on land values using machine learning algorithms and hedonic pricing methods. I predict agricultural land values with different machine learning algorithms, including ridge regression, least absolute shrinkage and selection operator, random forests, and extreme gradient boosting methods. To analyze the causal effects of water right seniority on agricultural land values, I use the double-selection LASSO technique. The second chapter presents the data used in the dissertation. A unique set of parcel sales from Property Valuation Division of Kansas constitute the backbone of the data used in the estimation. Along with parcel sales data, I collected detailed basis, water, tax, soil, weather, and urban influence data. This chapter provides detailed explanation of various data sources and

variable construction processes. The third chapter presents different machine learning models for irrigated agricultural land price predictions in Kansas. Researchers, and policymakers use different models and data sets for price prediction. Recently developed machine learning methods have the power to improve the predictive ability of the models estimated. In this chapter I estimate several machine learning models for predicting the agricultural land values in Kansas. Results indicate that the predictive power of the machine learning methods are stronger compared to standard econometric methods. Median absolute error in extreme gradient boosting estimation is 0.1312 whereas it is 0.6528 in simple OLS model. The fourth chapter examines whether water right seniority is capitalized into irrigated agricultural land values in Kansas. Using a unique data set of irrigated agricultural land sales, I analyze the causal effect of water right seniority on agricultural land values. A possible concern during the estimation of hedonic models is the omitted variable bias so we use double-selection LASSO regression and its variable selection properties to overcome the omitted variable bias. I also estimate generalized additive models to analyze the nonlinearities that may exist. Results show that water rights have a positive impact on irrigated land prices in Kansas. An additional year of water right seniority causes irrigated land value to increase nearly $17 per acre. Further analysis also suggest a nonlinear relationship between seniority and agricultural land prices.

The need to electronically store, manipulate and analyze large-scale, high-dimensional data sets requires new computational methods. This book presents new intelligent data management methods and tools, including new results from the field of inference. Leading experts also map out future directions of intelligent data analysis. This book will be a valuable reference for researchers exploring the interdisciplinary area between statistics and computer science as well as for professionals applying advanced data analysis methods in industry.

This dissertation focuses on using the machine learning technique, boosting, for causal inference in the instrumental variable (IV) regression models.

The SAGE Handbook of Research Methods in Political Science and International Relations offers a comprehensive overview of the field and its research processes through the empirical and research scholarship of leading international authors. The book is structured along the lines of applied research in the discipline: from formulating good research questions and designing a good research project, to various modes of theoretical argumentation, through conceptualization, to empirical measurement and analysis. Each chapter offers new approaches and builds upon existing methods. Through its seven parts, undergraduate and graduate students, researchers and practicing academics, will be guided through the design, methods and analysis of issues in Political Science and International Relations discipline: Part One: Formulating Good Research Questions and Designing Good Research Projects Part Two: Methods of Theoretical Argumentation Part Three: Conceptualization & Measurement Part Four: Large-Scale Data Collection & Representation Methods Part Five: Quantitative-Empirical Methods Part Six: Qualitative & “Mixed” Methods Part Seven: EITM & EMTI

Your hands-on reference guide to developing, training, and optimizing your machine learning models Key Features Your guide to learning efficient machine learning processes from scratch Explore expert techniques and hacks for a variety of machine learning concepts Write effective code in R, Python, Scala, and Spark to solve all your machine learning problems Book Description Machine learning makes it possible to learn about the unknowns and gain hidden insights into your datasets by mastering many tools and techniques. This book guides you to do just that in a very compact manner. After giving a quick overview of what machine learning is all about, Machine Learning Quick Reference jumps right into its core algorithms and demonstrates how they can be applied to real-world scenarios. From model evaluation to optimizing their performance, this book will introduce you to the best practices in machine learning. Furthermore, you will also look at the more advanced aspects such as training neural networks and work with different kinds of data, such as text, time-series, and sequential data. Advanced methods and techniques such as causal inference, deep Gaussian processes, and more are also covered. By the end of this book, you will be able to train fast, accurate machine learning models at your fingertips, which you can easily use as a point of reference. What you will learn Get a quick rundown of model selection, statistical modeling, and cross-validation Choose the best machine learning algorithm to solve your problem Explore kernel learning, neural networks, and time-series analysis Train deep learning models and optimize them for maximum performance Briefly cover Bayesian techniques and sentiment analysis in your NLP solution Implement probabilistic graphical models and causal inferences Measure and optimize the performance of your machine learning models Who this book is for If you're a machine learning practitioner, data scientist, machine learning developer, or engineer, this book will serve as a reference point in building machine learning solutions. You will also find this book useful if you're an intermediate machine learning developer or data scientist looking for a quick, handy reference to all the concepts of machine learning. You'll need some exposure to machine learning to get the best out of this book.

Causal inference -- the process of drawing a conclusion about the impact of an exposure on an outcome -- is foundational to biomedicine, where it is used to guide intervention. The current gold-standard approach for causal inference is randomized experimentation, such as randomized controlled trials (RCTs). Yet, randomized experiments, including RCTs, often enforce strict eligibility criteria that impede the generalizability of causal knowledge to the real world. Observational data, such as the electronic health record (EHR), is often regarded as a more representative source from which to generate causal knowledge. However, observational data is non-randomized, and therefore causal estimates from this source are susceptible to bias from confounders. This weakness complicates two central tasks of causal inference: the replication or evaluation of existing causal knowledge and the generation of new causal knowledge. In this dissertation I (i) address the feasibility of observational data to replicate existing causal knowledge and (ii) present new methods for the generation of causal knowledge with observational data, with a focus on the causal tasks of comparing an outcome between two cohorts and the estimation of attributable risks of exposures in a causal system.

This comprehensive encyclopedia, in A-Z format, provides easy access to relevant information for those seeking entry into any aspect within the broad field of Machine Learning. Most of the entries in this preeminent work include useful literature references.

Every day, decision-makers make choices among finite and discrete sets of alternatives. For example, people decide whether to walk, bike, take transit, or drive to work; shoppers decide which of the available brands of toothpaste to buy; and firms decide which vacant buildings they will rent for office space. Across these disparate domains, discrete choice models mathematically represent the procedures that analysts believe decision-makers are using to make such choices. Historically, the field of discrete choice modeling grew mainly out of economics, and this lineage has had long-lasting methodological ramifications. In particular, despite the great mathematical similarity between discrete choice models and models in statistics, machine learning, and causal inference, discrete choice research remains mostly siloed, seldom drawing from or contributing to methods in these related disciplines. In this dissertation, we help demolish the methodological silo around discrete choice re- search. Drawing from recent techniques in statistics, machine learning, and causal inference, we remove substantive limitations on the decision-making processes that could be be represented and predicted with previously available discrete choice methods. At the same time, by addressing concerns of discrete choice modelers, we make methodological contributions to the fields of statistics and machine learning, and we identify future research areas where discrete choice modelers are well suited to advancing the state of the art in causal inference. Importantly, the methodological advances described above were not divorced from to- day's societal concerns. Given that more and more government agencies are (unsuccessfully) attempting to raise bicycle commuting rates in their jurisdictions, we guide our interactions with the statistics, machine learning, and causal inference literatures by trying to more accurately model an individual's choice of commuting by bicycle. In particular, we use parametric link functions from statistics to better model the adoption and abandonment of bicycling. From machine learning, we use decision trees to represent the non-compensatory decision protocols that individuals appear to follow when deciding whether to commute by bicycle, and we use diagrams from the causal inference literature to gain insight into how we can bet- ter model the effects of bike lane investments on bicycle commute mode shares. All together, we not only make methodological contributions to the fields of discrete choice, statistics, machine learning, and causal inference, but we contribute to the efforts of transportation planners and modelers who are trying to make our cities and regions more sustainable and environmentally friendly. The methods developed in this dissertation have applications to strategic bicycle planning, helping analysts understand when certain interventions are not enough to cause people to abandon non-bicycle modes of travel at the desired rates and what alternative interventions might be more effective. In total, the specific contributions of this dissertation are the following: 1. We create a new spatial unit of analysis (the zone of likely travel) for the incorporation of roadway-level variables such as presence and type of bicycle infrastructure, roadway slopes, and traffic speeds into mode choice models. 2. We propose and demonstrate the novel use decision-tree methods for directly including roadway-level variables in mode choice models. 3. We create a new class of closed-form, finite-parameter, multinomial choice models that avoid an undesirable symmetry property that we describe in Chapter 3. 4. We create a procedure for using this new class of models to extend many existing binary choice models to the multinomial setting for the first time. 5. We create methods for creating new, symmetric and asymmetric, binary choice models. 6. We provide a microeconomic framework for interpreting decision trees by showing that decision trees represent a non-compensatory decision rule known as disjunctions-of-conjunctions and that such rules generalize many of the non-compensatory rules used in the discrete choice literature so far. 7. We propose and estimate the first bayesian model tree, thereby combining decision trees and discrete choice models in the first two-stage, semi-compensatory model that jointly: a) uses disjunctions-of-conjunctions for the choice-set generation stage, b) allows for context-dependent preference heterogeneity in the choice stage, and c) quantifies analyst uncertainty in the estimated disjunctions-of-conjunctions 8. We identify techniques such as the use of causal diagrams that can be borrowed from the causal inference literature to improve the ability of discrete choice modelers to predict outcomes under external changes or policy interventions such as investing in on-street bicycle lanes. 9. We identify areas of the causal inference literature that can be improved through the incorporation of techniques from discrete choice or through the application of causal inference techniques that are very relevant to discrete choice modellers yet only infrequently researched by traditional causal inference researchers. Through this dissertation, we empirically demonstrate most of our contributions using commute mode choice data from the San Francisco Bay Area. In every case, we found that the new models developed as part of this dissertation fit our data better than traditional discrete choice models. These results were stable across all measures of fit that were used, whether the measures were in-sample or out-of-sample, frequentist or bayesian. Beyond fit, all of our new models also proved to be qualitatively different than traditional discrete choice methods. Our new models provided insights and forecasts that both made more sense and were more accurate than their traditional counterparts. Finally, our contributions related to causal inference are the only items from the list above without empirical demonstrations. Instead, these contributions are bolstered by substantial literature review, discussion, and thought exercises that show the (general and bicycle specific) benefits of merging discrete choice and causal inference techniques.

Big Data in Omics and Imaging: Integrated Analysis and Causal Inference addresses the recent development of integrated genomic, epigenomic and imaging data analysis and causal inference in big data era. Despite significant progress in dissecting the genetic architecture of complex diseases by genome-wide association studies (GWAS), genome-wide expression studies (GWES), and epigenome-wide association studies (EWAS), the overall contribution of the new identified genetic variants is small and a large fraction of genetic variants is still hidden. Understanding the etiology and causal chain of mechanism underlying complex diseases remains elusive. It is time to bring big data, machine learning and causal revolution to developing a new generation of genetic analysis for shifting the current paradigm of genetic analysis from shallow association analysis to deep causal inference and from genetic analysis alone to integrated omics and imaging data analysis for unraveling the mechanism of complex diseases. FEATURES Provides a natural extension and companion volume to Big Data in Omic and Imaging: Association Analysis, but can be read independently. Introduce causal inference theory to genomic, epigenomic and imaging data analysis Develop novel statistics for genome-wide causation studies and epigenome-wide causation studies. Bridge the gap between the traditional association analysis and modern causation analysis Use combinatorial optimization methods and various causal models as a general framework for inferring multilevel omic and image causal networks Present statistical methods and computational algorithms for searching causal paths from genetic variant to disease Develop causal machine learning methods integrating causal inference and machine learning Develop statistics for testing significant difference in directed edge, path, and graphs, and for assessing causal relationships between two networks The book is designed for graduate students and researchers in genomics, epigenomics, medical image, bioinformatics, and data science. Topics covered are: mathematical formulation of causal inference, information geometry for causal inference, topology group and Haar measure, additive noise models, distance correlation, multivariate causal inference and causal networks, dynamic causal networks, multivariate and functional structural equation models, mixed structural equation models, causal inference with confounders, integer programming, deep learning and differential equations for wearable computing, genetic analysis of function-valued traits, RNA-seq data analysis, causal networks for genetic methylation analysis, gene expression and methylation deconvolution, cell –specific causal networks, deep learning for image segmentation and image analysis, imaging and genomic data analysis, integrated multilevel causal genomic, epigenomic and imaging data analysis.

Humes "Enquiry" gehört zu den wichtigsten Werken des großen schottischen Philosophen und ist einer der Klassiker der Philosophiegeschichte: Kant wurde nach seiner eigenen Aussage von diesem Werk aus seinem "dogmatischen Schlummer" geweckt und zu seiner kritischen Philosophie angeregt. In 12 Essays untersucht Hume verschiedenste Themen, die auch heute von Interesse sind, etwa Kausalität, den Skeptizismus, die Frage, ob wir frei oder determiniert sind oder ob es Wunder (bzw. letztlich: Religion) geben kann. Die zweisprachige Ausgabe gibt die vollständig durchgesehene klassische Übersetzung von Herbert Herring wieder und führt in Kommentar und Nachwort auf den neuesten Stand der Forschung. E-Book mit Seitenzählung der gedruckten Ausgabe: Buch und E-Book können parallel benutzt werden.

This hands-on book will help you make your machine learning models fairer, safer, and more reliable and in turn improve business outcomes. Every chapter introduces a new mission where you learn how to apply interpretation methods to realistic use cases with methods that work for any model type as well as methods specific for deep neural networks.

Fairness is a social norm and a legal requirement in today's society. Many laws and regulations (e.g., the Equal Credit Opportunity Act of 1974) have been established to prohibit discrimination and enforce fairness on several grounds, such as gender, age, sexual orientation, race, and religion, referred to as sensitive attributes. Nowadays machine learning algorithms are extensively applied to make important decisions in many real-world applications, e.g., employment, admission, and loans. Traditional machine learning algorithms aim to maximize predictive performance, e.g., accuracy. Consequently, certain groups may get unfairly treated when those algorithms are applied for decision-making. Therefore, it is an imperative task to develop fairness-aware machine learning algorithms such that the decisions made by them are not only accurate but also subject to fairness requirements. In the literature, machine learning researchers have proposed association-based fairness notions, e.g., statistical parity, disparate impact, equality of opportunity, etc., and developed respective discrimination mitigation approaches. However, these works did not consider that fairness should be treated as a causal relationship. Although it is well known that association does not imply causation, the gap between association and causation is not paid sufficient attention by the fairness researchers and stakeholders. The goal of this dissertation is to study fairness in machine learning, define appropriate fairness notions, and develop novel discrimination mitigation approaches from a causal perspective. Based on Pearl's structural causal model, we propose to formulate discrimination as causal effects of the sensitive attribute on the decision. We consider different types of causal effects to cope with different situations, including the path-specific effect for direct/indirect discrimination, the counterfactual effect for group/individual discrimination, and the path-specific counterfactual effect for general cases. In the attempt to measure discrimination, the unidentifiable situations pose an inevitable barrier to the accurate causal inference. To address this challenge, we propose novel bounding methods to accurately estimate the strength of unidentifiable fairness notions, including path-specific fairness, counterfactual fairness, and path-specific counterfactual fairness. Based on the estimation of fairness, we develop novel and efficient algorithms for learning fair classification models. Besides classification, we also investigate the discrimination issues in other machine learning scenarios, such as ranked data analysis.

The authors address the assumptions and methods that allow us to turn observations into causal knowledge, and use even incomplete causal knowledge in planning and prediction to influence and control our environment. What assumptions and methods allow us to turn observations into causal knowledge, and how can even incomplete causal knowledge be used in planning and prediction to influence and control our environment? In this book Peter Spirtes, Clark Glymour, and Richard Scheines address these questions using the formalism of Bayes networks, with results that have been applied in diverse areas of research in the social, behavioral, and physical sciences. The authors show that although experimental and observational study designs may not always permit the same inferences, they are subject to uniform principles. They axiomatize the connection between causal structure and probabilistic independence, explore several varieties of causal indistinguishability, formulate a theory of manipulation, and develop asymptotically reliable procedures for searching over equivalence classes of causal models, including models of categorical data and structural equation models with and without latent variables. The authors show that the relationship between causality and probability can also help to clarify such diverse topics in statistics as the comparative power of experimentation versus observation, Simpson's paradox, errors in regression models, retrospective versus prospective sampling, and variable selection. The second edition contains a new introduction and an extensive survey of advances and applications that have appeared since the first edition was published in 1993.

Machine Learning for Knowledge Discovery with R contains methodologies and examples for statistical modelling, inference, and prediction of data analysis. It includes many recent supervised and unsupervised machine learning methodologies such as recursive partitioning modelling, regularized regression, support vector machine, neural network, clustering, and causal-effect inference. Additionally, it emphasizes statistical thinking of data analysis, use of statistical graphs for data structure exploration, and result presentations. The book includes many real-world data examples from life-science, finance, etc. to illustrate the applications of the methods described therein. Key Features: Contains statistical theory for the most recent supervised and unsupervised machine learning methodologies. Emphasizes broad statistical thinking, judgment, graphical methods, and collaboration with subject-matter-experts in analysis, interpretation, and presentations. Written by statistical data analysis practitioner for practitioners. The book is suitable for upper-level-undergraduate or graduate-level data analysis course. It also serves as a useful desk-reference for data analysts in scientific research or industrial applications.

Advances in Machine Learning Research and Application: 2011 Edition is a ScholarlyEditions™ eBook that delivers timely, authoritative, and comprehensive information about Machine Learning. The editors have built Advances in Machine Learning Research and Application: 2011 Edition on the vast information databases of ScholarlyNews.™ You can expect the information about Machine Learning in this eBook to be deeper than what you can access anywhere else, as well as consistently reliable, authoritative, informed, and relevant. The content of Advances in Machine Learning Research and Application: 2011 Edition has been produced by the world’s leading scientists, engineers, analysts, research institutions, and companies. All of the content is from peer-reviewed sources, and all of it is written, assembled, and edited by the editors at ScholarlyEditions™ and available exclusively from us. You now have a source you can cite with authority, confidence, and credibility. More information is available at http://www.ScholarlyEditions.com/.

This book presents ground-breaking advances in the domain of causal structure learning. The problem of distinguishing cause from effect (“Does altitude cause a change in atmospheric pressure, or vice versa?”) is here cast as a binary classification problem, to be tackled by machine learning algorithms. Based on the results of the ChaLearn Cause-Effect Pairs Challenge, this book reveals that the joint distribution of two variables can be scrutinized by machine learning algorithms to reveal the possible existence of a “causal mechanism”, in the sense that the values of one variable may have been generated from the values of the other. This book provides both tutorial material on the state-of-the-art on cause-effect pairs and exposes the reader to more advanced material, with a collection of selected papers. Supplemental material includes videos, slides, and code which can be found on the workshop website. Discovering causal relationships from observational data will become increasingly important in data science with the increasing amount of available data, as a means of detecting potential triggers in epidemiology, social sciences, economy, biology, medicine, and other sciences.

The statistics profession is at a unique point in history. The need for valid statistical tools is greater than ever; data sets are massive, often measuring hundreds of thousands of measurements for a single subject. The field is ready to move towards clear objective benchmarks under which tools can be evaluated. Targeted learning allows (1) the full generalization and utilization of cross-validation as an estimator selection tool so that the subjective choices made by humans are now made by the machine, and (2) targeting the fitting of the probability distribution of the data toward the target parameter representing the scientific question of interest. This book is aimed at both statisticians and applied researchers interested in causal inference and general effect estimation for observational and experimental data. Part I is an accessible introduction to super learning and the targeted maximum likelihood estimator, including related concepts necessary to understand and apply these methods. Parts II-IX handle complex data structures and topics applied researchers will immediately recognize from their own research, including time-to-event outcomes, direct and indirect effects, positivity violations, case-control studies, censored data, longitudinal data, and genomic studies.