

# Table of Contents

Preface

xix

## Part 1: Causality – an Introduction

1

### Causality – Hey, We Have Machine Learning, So Why Even Bother? 3

A brief history of causality	4	A marketer's dilemma	9
Why causality? Ask babies!	5	Let's play doctor!	10
Interacting with the world	5	Associations in the wild	12
Confounding – relationships that are not real	6	Wrapping it up	12
How not to lose money... and human lives	9	References	12
		Join our book's Discord space	13

2

### Judea Pearl and the Ladder of Causation 15

From associations to logic and imagination – the Ladder of Causation	15	Computing counterfactuals	30
Associations	18	Time to code!	32
Let's practice!	20	Extra – is all machine learning causally the same?	33
What are interventions?	23	Causality and reinforcement learning	33
Changing the world	24	Causality and semi-supervised and unsupervised learning	34
Correlation and causation	26	Wrapping it up	34
What are counterfactuals?	28	References	35
Let's get weird (but formal)!	28		
The fundamental problem of causal inference	30		



## Regression, Observations, and Interventions

37

Starting simple – observational data and linear regression	37	Get involved!	48
Linear regression	37	To control or not to control?	48
p-values and statistical significance	41	Regression and structural models	49
Geometric interpretation of linear regression	42	SCMs	49
Reversing the order	42	Linear regression versus SCMs	49
Should we always control for all available covariates?	44	Finding the link	49
Navigating the maze	45	Regression and causal effects	51
If you don't know where you're going, you might end up somewhere else	45	Wrapping it up	53
		References	53

## Graphical Models

55

Graphs, graphs, graphs	55	Causal discovery	67
Types of graphs	56	Expert knowledge	67
Graph representations	58	Combining causal discovery and expert knowledge	67
Graphs in Python	60	Extra – is there causality beyond DAGs?	67
What is a graphical model?	63	Dynamical systems	67
DAG your pardon? Directed acyclic graphs in the causal wonderland	64	Cyclic SCMs	68
Definitions of causality	64	Wrapping it up	68
DAGs and causality	65	References	69
Let's get formal!	65		
Limitations of DAGs	66		
Sources of causal graphs in the real world	66		



## Forks, Chains, and Immoralities

71

Graphs and distributions and how to map between them 71

How to talk about independence 72

Choosing the right direction 73

Conditions and assumptions 74

Chains, forks, and colliders or... immoralities 78

A chain of events 78

Chains 79

Forks 80

Colliders, immoralities, or v-structures 82

Ambiguous cases 84

Forks, chains, colliders, and regression 85

Generating the chain dataset 87

Generating the fork dataset 88

Generating the collider dataset 89

Fitting the regression models 90

Wrapping it up 93

References 93

Join our book's Discord space 94

## Part 2: Causal Inference

### Nodes, Edges, and Statistical (In)dependence

97

You're gonna keep 'em d-separated 98

Practice makes perfect – d-separation 99

Estimand first! 102

We live in a world of estimators 102

So, what is an estimand? 102

The back-door criterion 104

What is the back-door criterion? 105

Back-door and equivalent estimands 105

The front-door criterion 107

Can GPS lead us astray? 108

London cabbies and the magic pebble 109

Opening the front door 110

Three simple steps toward the front door 111

Front-door in practice 112

Are there other criteria out there?

Let's do-calculus! 118

The three rules of do-calculus 119

Instrumental variables 120

Wrapping it up 122

Answer 122

References 123



## **The Four-Step Process of Causal Inference 125**

<b>Introduction to DoWhy and EconML</b>	<b>126</b>	<b>How to validate causal models</b>	<b>135</b>
Python causal ecosystem	126	Introduction to refutation tests	137
Why DoWhy?	128	<b>Full example</b>	<b>139</b>
Oui, mon ami, but what is DoWhy?	128	Step 1 – encode the assumptions	140
How about EconML?	129	Step 2 – getting the estimand	142
<b>Step 1 – modeling the problem</b>	<b>130</b>	Step 3 – estimate!	142
Creating the graph	130	Step 4 – refute them!	144
Building a CausalModel object	132	<b>Wrapping it up</b>	<b>149</b>
<b>Step 2 – identifying the estimand(s)</b>	<b>133</b>	<b>References</b>	<b>149</b>
<b>Step 3 – obtaining estimates</b>	<b>134</b>	<b>Join our book's Discord space</b>	<b>150</b>
<b>Step 4 – where's my validation set?</b>			
Refutation tests	135		

## **8**

## **Causal Models – Assumptions and Challenges 151**

<b>I am the king of the world! But am I?</b>	<b>152</b>	<b>...and more</b>	<b>162</b>
In between	152	Modularity	162
Identifiability	153	SUTVA	164
Lack of causal graphs	153	Consistency	164
Not enough data	154	<b>Call me names – spurious relationships in the wild</b>	<b>165</b>
Unverifiable assumptions	156	Names, names, names	165
An elephant in the room – hopeful or hopeless?	156	Should I ask you or someone who's not here?	166
Let's eat the elephant	156	DAG them!	166
<b>Positivity</b>	<b>157</b>	More selection bias	168
<b>Exchangeability</b>	<b>161</b>	<b>Wrapping it up</b>	<b>169</b>
Exchangeable subjects	161	<b>References</b>	<b>170</b>
Exchangeability versus confounding	161		



# Causal Inference and Machine Learning – from Matching to Meta-Learners

173

<b>The basics I – matching</b>	<b>174</b>	Jokes aside, say hi to the heterogeneous crowd	190
Types of matching	174	Waving the assumptions flag	192
Treatment effects – ATE versus ATT/ATC	175	You're the only one – modeling with S-Learner	192
Matching estimators	176	Small data	198
Implementing matching	178	S-Learner's vulnerabilities	199
<b>The basics II – propensity scores</b>	<b>183</b>	<b>T-Learner – together we can do more</b>	<b>200</b>
Matching in the wild	183	Forcing the split on treatment	200
Reducing the dimensionality with propensity scores	185	T-Learner in four steps and a formula	201
Propensity score matching (PSM)	185	Implementing T-Learner	202
<b>Inverse probability weighting (IPW)</b>	<b>186</b>	<b>X-Learner – a step further</b>	<b>204</b>
Many faces of propensity scores	186	Squeezing the lemon	204
Formalizing IPW	187	Reconstructing the X-Learner	205
Implementing IPW	187	X-Learner – an alternative formulation	207
IPW – practical considerations	188	Implementing X-Learner	208
<b>S-Learner – the Lone Ranger</b>	<b>188</b>	<b>Wrapping it up</b>	<b>212</b>
The devil's in the detail	189	<b>References</b>	<b>213</b>
Mom, Dad, meet CATE	190		

## 10

# Causal Inference and Machine Learning – Advanced Estimators, Experiments, Evaluations, and More

215

<b>Doubly robust methods – let's get more!</b>	<b>216</b>	DR-Learners – more options	224
Do we need another thing?	216	Targeted maximum likelihood estimator	224
Doubly robust is not equal to bulletproof...	218	<b>If machine learning is cool, how about double machine learning?</b>	<b>227</b>
...but it can bring a lot of value	218	Why DML and what's so double about it?	228
The secret doubly robust sauce	218	DML with DoWhy and EconML	231
Doubly robust estimator versus assumptions	220	Hyperparameter tuning with DoWhy and EconML	234
DR-Learner – crossing the chasm	220		



Is DML a golden bullet?	239	Kevin's challenge	252
Doubly robust versus DML	240	Opening the toolbox	253
What's in it for me?	241	Uplift models and performance	257
<b>Causal Forests and more</b>	<b>242</b>	Other metrics for continuous outcomes with multiple treatments	262
Causal trees	242	Confidence intervals	263
Forests overflow	242	Kevin's challenge's winning submission	264
Advantages of Causal Forests	242	When should we use CATE estimators for experimental data?	264
Causal Forest with DoWhy and EconML	243	Model selection – a simplified guide	265
<b>Heterogeneous treatment effects with experimental data – the uplift odyssey</b>	<b>245</b>	<b>Extra – counterfactual explanations</b>	<b>267</b>
The data	245	Bad faith or tech that does not know?	267
Choosing the framework	251	<b>Wrapping it up</b>	<b>268</b>
We don't know half of the story	251	<b>References</b>	<b>269</b>

# 11

## Causal Inference and Machine Learning – Deep Learning, NLP, and Beyond 273

<b>Going deeper – deep learning for heterogeneous treatment effects</b>	<b>274</b>	<b>econometrician goes Bayesian</b>	<b>297</b>
CATE goes deeper	274	Quasi-experiments	297
SNet	276	Twitter acquisition and our googling patterns	298
<b>Transformers and causal inference</b>	<b>284</b>	The logic of synthetic controls	298
The theory of meaning in five paragraphs	285	A visual introduction to the logic of synthetic controls	300
Making computers understand language	285	Starting with the data	302
From philosophy to Python code	286	Synthetic controls in code	303
LLMs and causality	286	Challenges	308
The three scenarios	288	<b>Wrapping it up</b>	<b>309</b>
CausalBert	292	<b>References</b>	<b>309</b>

## Causality and time series – when an



# Part 3: Causal Discovery

12

## Can I Have a Causal Graph, Please?

315

### Sources of causal knowledge

316

You and I, oversaturated

316

The power of a surprise

317

### Scientific insights

317

The logic of science

318

Hypotheses are a species

318

One logic, many ways

319

Controlled experiments

319

Randomized controlled trials (RCTs)

320

From experiments to graphs

321

Simulations

321

### Personal experience and domain knowledge

321

Personal experiences

322

Domain knowledge

323

### Causal structure learning

323

### Wrapping it up

324

References

324

Join our book's Discord space

326

13

## Causal Discovery and Machine Learning – from Assumptions to Applications

327

### Causal discovery – assumptions refresher

328

Gearing up

328

Always trying to be faithful...

328

...but it's difficult sometimes

328

Minimalism is a virtue

329

### The four (and a half) families

329

The four streams

329

### Introduction to gCastle

331

Hello, gCastle!

331

Synthetic data in gCastle

331

Fitting your first causal discovery model

336

Visualizing the model

336

Model evaluation metrics

338

### Constraint-based causal discovery

341

Constraints and independence

341

Leveraging the independence structure to recover the graph

342

PC algorithm – hidden challenges

345

PC algorithm for categorical data

346

### Score-based causal discovery

347

Tabula rasa – starting fresh

347

GES – scoring

347

GES in gCastle

348



<b>Functional causal discovery</b>	<b>349</b>	<b>GOLEMs don't cry</b>	<b>363</b>
The blessings of asymmetry	349	The comparison	363
ANM model	350	<b>Encoding expert knowledge</b>	<b>366</b>
Assessing independence	353	What is expert knowledge?	366
LiNGAM time	355	Expert knowledge in gCastle	366
<b>Gradient-based causal discovery</b>	<b>360</b>	<b>Wrapping it up</b>	<b>368</b>
What exactly is so gradient about you?	360	<b>References</b>	<b>368</b>
Shed no tears	362		

## 14

### **Causal Discovery and Machine Learning – Advanced Deep Learning and Beyond** **371**

<b>Advanced causal discovery with deep learning</b>	<b>372</b>	<b>Other approaches to confounded data</b>	<b>392</b>
From generative models to causality	372	<b>Extra – going beyond observations</b>	<b>393</b>
Looking back to learn who you are	373	ENCO	393
DECI's internal building blocks	373	ABCI	393
DECI in code	375	<b>Causal discovery – real-world applications, challenges, and open problems</b>	<b>394</b>
DECI is end-to-end	387	<b>Wrapping it up!</b>	<b>395</b>
<b>Causal discovery under hidden confounding</b>	<b>387</b>	<b>References</b>	<b>396</b>
The FCI algorithm	388		

## 15

### **Epilogue** **399**

<b>What we've learned in this book</b>	<b>399</b>	<b>Causality and business</b>	<b>403</b>
<b>Five steps to get the best out of your causal project</b>	<b>400</b>	How causal doers go from vision to implementation	403
Starting with a question	400	<b>Toward the future of causal ML</b>	<b>405</b>
Obtaining expert knowledge	401	Where are we now and where are we heading?	406
Generating hypothetical graph(s)	401	Causal benchmarks	406
Check identifiability	402	Causal data fusion	407
Falsifying hypotheses	402	Intervening agents	407



Causal structure learning	408	Let's stay in touch	410
Imitation learning	408	Wrapping it up	411
Learning causality	409	References	411
		Join our book's Discord space	412
<b>Index</b>			<b>413</b>
<b>Other Books You May Enjoy</b>			<b>426</b>