

Machine Learning and Data Science

Table of Contents

Introduction – Machine Learning and Data Science.....	9
What is Big Data?.....	9
What do we do with Big Data?.....	11
New Econometric Methods	11
Challenges of Machine Learning?	11
Computational Engines.....	12
 Chapter 1 – Core Concepts in Statistical Learning	 15
Modeling.....	15
Statistical Approach	15
Machine Learning Approach.....	16
Supervised Learning.....	16
Bias-Variance Tradeoff.....	17
Cross Validation	18
The Bootstrap	19
Computational Exercise.....	21
Further Reading	21
 Chapter 2 – Regression and Prediction.....	 23
Linear Regression	23
Why linear regression?.....	24
When the number of parameters is large.....	24
Model Selection.....	24
Subset Selection	24
Shrinkage.....	25
Elastic-Net.....	28
The Oracle Property	29
Spike and Slab.....	29
Hedonic Modeling	31

Computational Exercise.....	33
Further Reading	33
Chapter 3 – Classification	35
Classification in Economics	35
Model Choices for Classification	36
Random Utility Models	37
Assessing accuracy	38
Choice Based Sampling.....	38
Classification in Panel Data with Neural networks	39
Model Framework	39
Simulation Study (Parret, 2017).....	40
Computational Exercise.....	42
Further Reading	42
Chapter 4 – Unsupervised Learning	43
What is Unsupervised Learning	43
Dimensionality Reduction	43
Principle Components	43
Why is PCA linear.....	44
PCA in action	44
PCA Regression	45
Factor Models.....	46
Selecting the Number of Components.....	46
Clustering	47
K-Means Clustering	47
Hierarchical Clustering	47
Clustering in action (Harding Lovenheim, 2017)	48
Computational Exercise.....	50
Further Reading	50

Chapter 5 – Causal Inference	51
Rubin Causal Model	51
Potential Outcomes	51
Measuring Treatment	52
Randomized Control Trial (RCT).....	53
Example: Time of Use Experiment.....	53
Model Specification	54
Selection Bias.....	54
Addressing Selection Bias.....	55
Difference-in-Differences.....	56
Matching Estimation.....	56
Identification.....	58
Computational Exercise.....	59
Further Reading	59
 Chapter 6 – From Prediction to Causal Inference.....	 61
Is Prediction Enough	61
Structural Models and Treatment Effects	61
Many Instruments	62
Selection Among Controls.....	62
Double Machine Learning (Chernozhukov et. al. 2017)	63
Double ML Example: 401(k) Pension Plan.....	65
Computational Exercise.....	66
Further Reading	66
 Chapter 7 – Personalization and Heterogeneous Treatment Effects	 67
Quantile Regression.....	67
Quantiles for Count Data.....	67
Model Specification	68
Model Formulation.....	68
Estimation	68
Heterogeneity and Trees.....	69
Complexity Control.....	71

Bagging.....	72
Random Forests	72
Human decisions and machine predictions (Kleinberg et. al., 2017)	72
Heterogeneous Treatment Effects (Athey and Wager, 2017).....	74
Computational Exercise.....	75
Further Reading	75
Chapter 8 – Deep Learning.....	77
What are Neural Networks	77
Why Neural Networks Matter	77
Autoencoders	78
Further Reading	80
Appendix.....	81
PowerPoint Presentation	177