
Contents

Preface	xiii
I Introduction	1
1 Notations and data	3
1.1 Notations	3
1.2 Dataset	4
2 Introduction	9
2.1 Context	9
2.2 Portfolio construction: the workflow	10
2.3 Machine learning is no magic wand	11
3 Factor investing and asset pricing anomalies	13
3.1 Introduction	14
3.2 Detecting anomalies	15
3.2.1 Challenges	15
3.2.2 Simple portfolio sorts	15
3.2.3 Factors	17
3.2.4 Fama-Macbeth regressions	22
3.2.5 Factor competition	25
3.2.6 Advanced techniques	26
3.3 Factors or characteristics?	27
3.4 Hot topics: momentum, timing and ESG	28
3.4.1 Factor momentum	28
3.4.2 Factor timing	29
3.4.3 The green factors	30
3.5 The links with machine learning	30
3.5.1 A short list of recent references	31
3.5.2 Explicit connections with asset pricing models	31
3.6 Coding exercises	34
4 Data preprocessing	35
4.1 Know your data	35
4.2 Missing data	38
4.3 Outlier detection	40
4.4 Feature engineering	41
4.4.1 Feature selection	41
4.4.2 Scaling the predictors	41
4.5 Labelling	42
4.5.1 Simple labels	42
4.5.2 Categorical labels	43

4.5.3	The triple barrier method	44
4.5.4	Filtering the sample	45
4.5.5	Return horizons	46
4.6	Handling persistence	47
4.7	Extensions	47
4.7.1	Transforming features	47
4.7.2	Macro-economic variables	48
4.7.3	Active learning	48
4.8	Additional code and results	50
4.8.1	Impact of rescaling: graphical representation	50
4.8.2	Impact of rescaling: toy example	52
4.9	Coding exercises	53
II	Common supervised algorithms	55
5	Penalized regressions and sparse hedging for minimum variance portfolios	57
5.1	Penalized regressions	57
5.1.1	Simple regressions	57
5.1.2	Forms of penalizations	58
5.1.3	Illustrations	60
5.2	Sparse hedging for minimum variance portfolios	62
5.2.1	Presentation and derivations	62
5.2.2	Example	65
5.3	Predictive regressions	67
5.3.1	Literature review and principle	67
5.3.2	Code and results	68
5.4	Coding exercise	68
6	Tree-based methods	69
6.1	Simple trees	69
6.1.1	Principle	69
6.1.2	Further details on classification	71
6.1.3	Pruning criteria	72
6.1.4	Code and interpretation	73
6.2	Random forests	76
6.2.1	Principle	76
6.2.2	Code and results	78
6.3	Boosted trees: Adaboost	79
6.3.1	Methodology	79
6.3.2	Illustration	82
6.4	Boosted trees: extreme gradient boosting	82
6.4.1	Managing loss	83
6.4.2	Penalization	83
6.4.3	Aggregation	84
6.4.4	Tree structure	85
6.4.5	Extensions	86
6.4.6	Code and results	86
6.4.7	Instance weighting	88
6.5	Discussion	89
6.6	Coding exercises	90

7	Neural networks	91
7.1	The original perceptron	92
7.2	Multilayer perceptron	93
7.2.1	Introduction and notations	93
7.2.2	Universal approximation	96
7.2.3	Learning via back-propagation	97
7.2.4	Further details on classification	100
7.3	How deep we should go and other practical issues	101
7.3.1	Architectural choices	101
7.3.2	Frequency of weight updates and learning duration	102
7.3.3	Penalizations and dropout	103
7.4	Code samples and comments for vanilla MLP	104
7.4.1	Regression example	104
7.4.2	Classification example	107
7.4.3	Custom losses	111
7.5	Recurrent networks	112
7.5.1	Presentation	112
7.5.2	Code and results	114
7.6	Other common architectures	117
7.6.1	Generative adversarial networks	117
7.6.2	Autoencoders	118
7.6.3	A word on convolutional networks	119
7.6.4	Advanced architectures	121
7.7	Coding exercise	121
8	Support vector machines	123
8.1	SVM for classification	123
8.2	SVM for regression	126
8.3	Practice	127
8.4	Coding exercises	128
9	Bayesian methods	129
9.1	The Bayesian framework	129
9.2	Bayesian sampling	131
9.2.1	Gibbs sampling	131
9.2.2	Metropolis-Hastings sampling	131
9.3	Bayesian linear regression	132
9.4	Naive Bayes classifier	135
9.5	Bayesian additive trees	138
9.5.1	General formulation	138
9.5.2	Priors	138
9.5.3	Sampling and predictions	139
9.5.4	Code	141
III	From predictions to portfolios	143
10	Validating and tuning	145
10.1	Learning metrics	145
10.1.1	Regression analysis	145
10.1.2	Classification analysis	147
10.2	Validation	151

10.2.1	The variance-bias tradeoff: theory	151
10.2.2	The variance-bias tradeoff: illustration	154
10.2.3	The risk of overfitting: principle	156
10.2.4	The risk of overfitting: some solutions	157
10.3	The search for good hyperparameters	158
10.3.1	Methods	158
10.3.2	Example: grid search	160
10.3.3	Example: Bayesian optimization	162
10.4	Short discussion on validation in backtests	163
11	Ensemble models	165
11.1	Linear ensembles	166
11.1.1	Principles	166
11.1.2	Example	167
11.2	Stacked ensembles	170
11.2.1	Two-stage training	170
11.2.2	Code and results	170
11.3	Extensions	172
11.3.1	Exogenous variables	172
11.3.2	Shrinking inter-model correlations	173
11.4	Exercise	176
12	Portfolio backtesting	177
12.1	Setting the protocol	177
12.2	Turning signals into portfolio weights	179
12.3	Performance metrics	181
12.3.1	Discussion	181
12.3.2	Pure performance and risk indicators	182
12.3.3	Factor-based evaluation	183
12.3.4	Risk-adjusted measures	184
12.3.5	Transaction costs and turnover	184
12.4	Common errors and issues	185
12.4.1	Forward looking data	185
12.4.2	Backtest overfitting	185
12.4.3	Simple safeguards	187
12.5	Implication of non-stationarity: forecasting is hard	187
12.5.1	General comments	187
12.5.2	The no free lunch theorem	188
12.6	First example: a complete backtest	189
12.7	Second example: backtest overfitting	193
12.8	Coding exercises	196
IV	Further important topics	197
13	Interpretability	199
13.1	Global interpretations	200
13.1.1	Simple models as surrogates	200
13.1.2	Variable importance (tree-based)	201
13.1.3	Variable importance (agnostic)	203
13.1.4	Partial dependence plot	205
13.2	Local interpretations	206

18.8	Chapter 13	281
18.8.1	EW portfolios with the tidyverse	281
18.8.2	Advanced weighting function	282
18.8.3	Functional programming in the backtest	283
18.9	Chapter 16	285
18.10	Chapter 17	285
	Bibliography	289
	Index	319