

Contents

Preface	<i>Page xv</i>
Notation Used	xviii
List of Abbreviations	xix
1 Introduction	1
1.1 Statistics and Machine Learning \boxed{A}^{\odot}	1
1.2 Environmental Data Science \boxed{A}^{\odot}	6
1.3 A Simple Example of Curve Fitting \boxed{A}^{\odot}	9
1.4 Main Types of Data Problems \boxed{A}^{\odot}	12
1.4.1 Supervised Learning \boxed{A}^{\odot}	15
1.4.2 Unsupervised Learning \boxed{A}^{\odot}	16
1.4.3 Reinforced Learning \boxed{A}^{\odot}	17
1.5 Curse of Dimensionality \boxed{A}^{\odot}	17
2 Basics	19
2.1 Data Types \boxed{A}^{\odot}	19
2.2 Probability \boxed{A}^{\odot}	19
2.3 Probability Density \boxed{A}^{\odot}	22
2.4 Expectation and Mean \boxed{A}^{\odot}	24
2.5 Variance and Standard Deviation \boxed{A}^{\odot}	25
2.6 Covariance \boxed{A}^{\odot}	26
2.7 Online Algorithms for Mean, Variance and Covariance \boxed{C}^{\odot}	27
2.8 Median and Median Absolute Deviation \boxed{A}^{\odot}	29
2.9 Quantiles \boxed{A}^{\odot}	30
2.10 Skewness and Kurtosis \boxed{B}^{\odot}	32
2.11 Correlation \boxed{A}^{\odot}	33
2.11.1 Pearson Correlation \boxed{A}^{\odot}	33
2.11.2 Serial Correlation \boxed{A}^{\odot}	36
2.11.3 Spearman Rank Correlation \boxed{A}^{\odot}	38
2.11.4 Kendall Rank Correlation \boxed{A}^{\odot}	39
2.11.5 Biweight Midcorrelation \boxed{B}^{\odot}	40
2.12 Exploratory Data Analysis \boxed{A}^{\odot}	41
2.12.1 Histograms \boxed{A}^{\odot}	41
2.12.2 Quantile–Quantile (Q–Q) Plots \boxed{B}^{\odot}	42

2.12.3	Boxplots [A]⊙	45
2.13	Mahalanobis Distance [A]⊙	47
2.13.1	Mahalanobis Distance and Principal Component Analysis [B]⊙	47
2.14	Bayes Theorem [A]⊙	49
2.15	Classification [A]⊙	52
2.16	Clustering [A]⊙	54
2.17	Information Theory [B]⊙	56
2.17.1	Entropy [B]⊙	58
2.17.2	Joint Entropy and Conditional Entropy [B]⊙	59
2.17.3	Relative Entropy [B]⊙	60
2.17.4	Mutual Information [B]⊙	61
	Exercises	62
3	Probability Distributions	65
3.1	Binomial Distribution [A]⊙	66
3.2	Poisson Distribution [B]⊙	68
3.3	Multinomial Distribution [B]⊙	68
3.4	Gaussian Distribution [A]⊙	69
3.5	Maximum Likelihood Estimation [A]⊙	73
3.6	Multivariate Gaussian Distribution [B]⊙	75
3.7	Conditional and Marginal Gaussian Distributions [C]⊙	77
3.8	Gamma Distribution [B]⊙	78
3.9	Beta Distribution [B]⊙	80
3.10	Von Mises Distribution [C]⊙	82
3.11	Extreme Value Distributions [C]⊙	83
3.12	Gaussian Mixture Model [B]⊙	86
3.12.1	Expectation–Maximization (EM) Algorithm [B]⊙	90
3.13	Kernel Density Estimation [B]⊙	91
3.14	Re-expressing Data [A]⊙	93
3.15	Student <i>t</i> -distribution [B]⊙	95
3.16	Chi-squared Distribution [B]⊙	97
	Exercises	99
4	Statistical Inference	101
4.1	Hypothesis Testing [A]⊙	101
4.2	Student <i>t</i> -test [A]⊙	104
4.2.1	One-Sample <i>t</i> -test [A]⊙	105
4.2.2	Independent Two-Sample <i>t</i> -test [A]⊙	105
4.2.3	Dependent <i>t</i> -test for Paired Samples [A]⊙	107
4.2.4	Serial Correlation [A]⊙	107
4.2.5	Significance Test for Correlation [A]⊙	109
4.3	Non-parametric Alternatives to <i>t</i> -test [B]⊙	111
4.3.1	Wilcoxon–Mann–Whitney Test [C]⊙	111
4.3.2	Wilcoxon Signed-Rank Test [C]⊙	114
4.4	Confidence Interval [A]⊙	115

4.4.1	Confidence Interval for Population Mean	B ☺	116
4.4.2	Confidence Interval for Correlation	B ☺	118
4.5	Goodness-of-Fit Tests	B ☺	119
4.5.1	One-Sample Goodness-of-Fit Tests	B ☺	119
4.5.2	Two-Sample Goodness-of-Fit Tests	B ☺	121
4.6	Test of Variances	B ☺	124
4.7	Mann–Kendall Trend Test	B ☺	125
4.8	Bootstrapping	A ☺	126
4.9	Field Significance	B ☺	131
	Exercises		134
5	Linear Regression		137
5.1	Simple Linear Regression	A ☺	137
5.1.1	Partition of Sums of Squares	A ☺	139
5.1.2	Confidence Interval for Regression Parameters	B ☺	141
5.1.3	Confidence Interval and Prediction Interval for the Response Variable	B ☺	143
5.1.4	Serial Correlation	B ☺	144
5.2	Multiple Linear Regression	A ☺	145
5.2.1	Gauss–Markov Theorem	B ☺	147
5.2.2	Partition of Sums of Squares	B ☺	148
5.2.3	Standardized Predictors	A ☺	148
5.2.4	Analysis of Variance (ANOVA)	B ☺	149
5.2.5	Confidence and Prediction Intervals	B ☺	151
5.3	Multivariate Linear Regression	B ☺	153
5.4	Online Learning with Linear Regression	C ☺	154
5.5	Circular and Categorical Data	A ☺	156
5.6	Predictor Selection	C ☺	158
5.7	Ridge Regression	B ☺	161
5.8	Lasso	C ☺	165
5.9	Quantile Regression	C ☺	167
5.10	Generalized Least Squares	C ☺	168
5.10.1	Optimal Fingerprinting in Climate Change	C ☺	170
	Exercises		170
6	Neural Networks		173
6.1	McCulloch and Pitts Model	B ☺	174
6.2	Perceptrons	A ☺	175
6.2.1	Limitation of Perceptrons	B ☺	178
6.3	Multi-layer Perceptrons	A ☺	180
6.3.1	Comparison with Polynomials	B ☺	185
6.3.2	Hidden Neurons	A ☺	186
6.3.3	Monotonic Multi-layer Perceptron Model	C ☺	187
6.4	Extreme Learning Machines	A ☺	189
6.4.1	Online Learning	B ☺	193
6.4.2	Random Vector Functional Link	C ☺	194

6.5	Radial Basis Functions	\boxed{B}^{\odot}	195
6.6	Modelling Conditional Distributions	\boxed{B}^{\odot}	199
6.6.1	Mixture Density Network	\boxed{B}^{\odot}	201
6.7	Quantile Regression	\boxed{C}^{\odot}	204
6.8	Historical Development of NN in Environmental Science	\boxed{B}^{\odot}	207
6.8.1	Remote Sensing	\boxed{B}^{\odot}	208
6.8.2	Hydrology	\boxed{B}^{\odot}	212
6.8.3	Atmospheric Science	\boxed{B}^{\odot}	213
6.8.4	Oceanography	\boxed{B}^{\odot}	213
	Exercises		214
7	Non-linear Optimization		216
7.1	Extrema and Saddle Points	\boxed{A}^{\odot}	216
7.2	Gradient Vector in Optimization	\boxed{A}^{\odot}	219
7.3	Back-Propagation	\boxed{B}^{\odot}	221
7.4	Training Protocol	\boxed{A}^{\odot}	224
7.5	Gradient Descent Method	\boxed{A}^{\odot}	225
7.6	Stochastic Gradient Descent	\boxed{B}^{\odot}	227
7.7	Conjugate Gradient Method	\boxed{C}^{\odot}	229
7.8	Quasi-Newton Methods	\boxed{C}^{\odot}	232
7.9	Non-linear Least Squares Methods	\boxed{C}^{\odot}	234
7.10	Evolutionary Algorithms	\boxed{B}^{\odot}	236
7.11	Hill Climbing	\boxed{C}^{\odot}	238
7.12	Genetic Algorithm	\boxed{C}^{\odot}	239
7.13	Differential Evolution	\boxed{C}^{\odot}	241
	Exercises		244
8	Learning and Generalization		245
8.1	Mean Squared Error and Maximum Likelihood	\boxed{A}^{\odot}	245
8.2	Objective Functions and Robustness	\boxed{A}^{\odot}	247
8.3	Variance and Bias Errors	\boxed{A}^{\odot}	250
8.4	Regularization	\boxed{A}^{\odot}	251
8.4.1	Weight Penalty	\boxed{A}^{\odot}	251
8.4.2	Early Stopping	\boxed{A}^{\odot}	254
8.5	Cross-Validation	\boxed{A}^{\odot}	255
8.6	Hyperparameter Tuning	\boxed{A}^{\odot}	258
8.7	Ensemble Methods	\boxed{A}^{\odot}	261
8.7.1	Bagging	\boxed{A}^{\odot}	263
8.7.2	Error of Ensemble	\boxed{B}^{\odot}	263
8.7.3	Unequal Weighting of Ensemble Members	\boxed{C}^{\odot}	266
8.7.4	Stacking	\boxed{C}^{\odot}	268
8.8	Dropout	\boxed{B}^{\odot}	269
8.9	Maximum Norm Constraint	\boxed{B}^{\odot}	271
8.10	Bayesian Model Selection	\boxed{B}^{\odot}	272
8.11	Information Criterion	\boxed{B}^{\odot}	273
8.11.1	Bayesian Information Criterion	\boxed{C}^{\odot}	273

<i>Contents</i>	ix
8.11.2 Akaike Information Criterion [C]☺	275
8.12 Bayesian Model Averaging [C]☺	278
8.13 Interpretable ML [C]☺	280
Exercises	281
9 Principal Components and Canonical Correlation	283
9.1 Principal Component Analysis (PCA) [A]☺	283
9.1.1 Geometric Approach to PCA [A]☺	284
9.1.2 Eigenvector Approach to PCA [A]☺	284
9.1.3 Real and Complex Data [C]☺	288
9.1.4 Orthogonality Relations [A]☺	289
9.1.5 PCA of the Tropical Pacific Climate Variability [A]☺	290
9.1.6 Scaling the PCs and Eigenvectors [A]☺	298
9.1.7 Degeneracy of Eigenvalues [A]☺	299
9.1.8 A Smaller Covariance Matrix [A]☺	299
9.1.9 How Many Modes to Retain [B]☺	300
9.1.10 Temporal and Spatial Mean Removal [B]☺	302
9.1.11 Singular Value Decomposition [B]☺	302
9.1.12 Missing Data [C]☺	304
9.2 Rotated PCA [B]☺	305
9.2.1 E-frame Rotation [B]☺	308
9.2.2 A-frame Rotation [B]☺	311
9.2.3 Advantages and Disadvantages of Rotation [A]☺	315
9.3 PCA for Two-Dimensional Vectors [C]☺	317
9.4 Canonical Correlation Analysis (CCA) [B]☺	320
9.4.1 CCA Theory [B]☺	320
9.4.2 Pre-filter with PCA [B]☺	324
9.5 Maximum Covariance Analysis [B]☺	326
Exercises	327
10 Unsupervised Learning	330
10.1 Clustering [B]☺	330
10.1.1 Distance Measure [B]☺	331
10.1.2 Model Evaluation [B]☺	332
10.2 Non-hierarchical Clustering [B]☺	337
10.2.1 K-means Clustering [B]☺	337
10.2.2 Nucleated Agglomerative Clustering [C]☺	338
10.2.3 Gaussian Mixture Model [C]☺	339
10.3 Hierarchical Clustering [C]☺	339
10.4 Self-Organizing Map [C]☺	343
10.4.1 Applications of SOM [C]☺	345
10.5 Autoencoder [A]☺	347
10.6 Non-linear Principal Component Analysis [B]☺	349
10.6.1 Overfitting [C]☺	358
10.6.2 Closed Curves [C]☺	360
10.7 Other Non-linear Dimensionality Reduction Methods [C]☺	363

10.8 Non-linear Canonical Correlation Analysis [C]⊙	365
Exercises	369
11 Time Series	372
11.1 Fourier Analysis [A]⊙	372
11.1.1 Fourier Series [A]⊙	373
11.1.2 Discrete Fourier Transform [A]⊙	374
11.1.3 Continuous-Time Fourier Transform [B]⊙	375
11.1.4 Discrete-Time Fourier Transform [B]⊙	376
11.2 Windows [A]⊙	376
11.3 Spectrum [A]⊙	379
11.3.1 Effects of Window Functions [A]⊙	380
11.3.2 Trend Removal [A]⊙	382
11.3.3 Nyquist Frequency and Aliasing [A]⊙	382
11.3.4 Smoothing the Spectrum [A]⊙	384
11.3.5 Confidence Interval [B]⊙	385
11.3.6 Examples [B]⊙	385
11.3.7 Fast Fourier Transform [B]⊙	386
11.3.8 Relation with Auto-covariance [B]⊙	387
11.3.9 Rotary Spectrum for 2-D Vectors [C]⊙	387
11.3.10 Wavelets [C]⊙	390
11.4 Cross-Spectrum [B]⊙	393
11.5 Filtering [A]⊙	395
11.5.1 Periodic Signals [A]⊙	396
11.5.2 Ideal Filters [A]⊙	397
11.5.3 Finite Impulse Response Filters [B]⊙	398
11.6 Averaging [A]⊙	401
11.6.1 Moving Average Filters [A]⊙	402
11.6.2 Grid-Scale Noise [B]⊙	402
11.6.3 Linearization from Time-Averaging [B]⊙	404
11.7 Singular Spectrum Analysis [B]⊙	405
11.7.1 Multichannel Singular Spectrum Analysis [B]⊙	407
11.8 Auto-regressive Process [B]⊙	410
11.8.1 AR(p) Process [B]⊙	411
11.8.2 AR(1) Process [B]⊙	412
11.8.3 AR(2) Process [C]⊙	413
11.9 Box–Jenkins Models [C]⊙	414
11.9.1 Moving Average (MA) Process [C]⊙	414
11.9.2 Auto-regressive Moving Average (ARMA) Model [C]⊙	414
11.9.3 Auto-regressive Integrated Moving Average (ARIMA) Model [C]⊙	415
Exercises	416

<i>Contents</i>	xi
12 Classification	418
12.1 Linear Discriminant Analysis $\boxed{\text{A}}^{\ominus}$	419
12.1.1 Fisher Linear Discriminant $\boxed{\text{B}}^{\ominus}$	421
12.2 Logistic Regression $\boxed{\text{A}}^{\ominus}$	424
12.2.1 Multiclass Logistic Regression $\boxed{\text{B}}^{\ominus}$	425
12.3 Naive Bayes Classifier $\boxed{\text{B}}^{\ominus}$	427
12.4 K -nearest Neighbours $\boxed{\text{B}}^{\ominus}$	428
12.5 Extreme Learning Machine Classifier $\boxed{\text{A}}^{\ominus}$	430
12.6 Cross-Entropy $\boxed{\text{A}}^{\ominus}$	432
12.7 Multi-layer Perceptron Classifier $\boxed{\text{A}}^{\ominus}$	434
12.8 Class Imbalance $\boxed{\text{A}}^{\ominus}$	436
Exercises	438
13 Kernel Methods	440
13.1 From Neural Networks to Kernel Methods $\boxed{\text{B}}^{\ominus}$	441
13.2 Primal and Dual Solutions for Linear Regression $\boxed{\text{B}}^{\ominus}$	442
13.3 Kernels $\boxed{\text{B}}^{\ominus}$	444
13.4 Kernel Ridge Regression $\boxed{\text{B}}^{\ominus}$	448
13.5 Advantages and Disadvantages $\boxed{\text{B}}^{\ominus}$	449
13.6 Pre-image Problem $\boxed{\text{C}}^{\ominus}$	451
13.7 Support Vector Machines (SVM) $\boxed{\text{B}}^{\ominus}$	453
13.7.1 Linearly Separable Case $\boxed{\text{B}}^{\ominus}$	454
13.7.2 Linearly Non-separable Case $\boxed{\text{B}}^{\ominus}$	458
13.7.3 Non-linear Classification by SVM $\boxed{\text{B}}^{\ominus}$	460
13.7.4 Multi-class Classification by SVM $\boxed{\text{C}}^{\ominus}$	461
13.7.5 Support Vector Regression $\boxed{\text{C}}^{\ominus}$	462
13.8 Gaussian Processes $\boxed{\text{B}}^{\ominus}$	463
13.8.1 Learning the Hyperparameters $\boxed{\text{B}}^{\ominus}$	466
13.8.2 Other Common Kernels $\boxed{\text{C}}^{\ominus}$	467
13.9 Kernel Principal Component Analysis $\boxed{\text{C}}^{\ominus}$	469
Exercises	471
14 Decision Trees, Random Forests and Boosting	473
14.1 Classification and Regression Trees (CART) $\boxed{\text{A}}^{\ominus}$	474
14.1.1 Relative Importance of Predictors $\boxed{\text{B}}^{\ominus}$	479
14.1.2 Surrogate Splits $\boxed{\text{B}}^{\ominus}$	480
14.2 Random Forests $\boxed{\text{A}}^{\ominus}$	482
14.2.1 Extremely Randomized Trees (Extra Trees) $\boxed{\text{B}}^{\ominus}$	487
14.3 Boosting $\boxed{\text{A}}^{\ominus}$	487
14.3.1 Gradient Boosting $\boxed{\text{B}}^{\ominus}$	488
Exercises	492

15 Deep Learning	494
15.1 Transfer Learning A ☺	498
15.2 Convolutional Neural Network B ☺	499
15.2.1 Convolution Operation B ☺	499
15.2.2 Pooling B ☺	502
15.2.3 AlexNet B ☺	503
15.2.4 Residual Neural Network (ResNet) B ☺	505
15.2.5 Data Augmentation C ☺	506
15.2.6 Applications in Environment Science B ☺	506
15.3 Encoder–Decoder Network B ☺	507
15.3.1 U-net B ☺	508
15.4 Time Series C ☺	510
15.4.1 Long Short-Term Memory (LSTM) Network C ☺	510
15.4.2 Temporal Convolutional Network C ☺	513
15.5 Generative Adversarial Network C ☺	514
Exercises	517
16 Forecast Verification and Post-processing	518
16.1 Binary Classes B ☺	519
16.1.1 Skill Scores for Binary Classes B ☺	524
16.2 Multiple Classes C ☺	527
16.3 Probabilistic Forecasts for Binary Classes B ☺	528
16.3.1 Reliability Diagram B ☺	529
16.4 Probabilistic Forecasts for Multiple Classes B ☺	531
16.5 Continuous Variables B ☺	532
16.5.1 Forecast Scores B ☺	532
16.5.2 Skill Scores B ☺	533
16.6 Probabilistic Forecasts for Continuous Variables B ☺	535
16.7 Minimizing Loss C ☺	535
16.8 Spurious Skills A ☺	536
16.9 Extrapolation A ☺	538
16.10 Post-processing C ☺	541
16.11 Downscaling C ☺	543
16.11.1 Reduced Variance C ☺	546
Exercises	547
17 Merging of Machine Learning and Physics	549
17.1 Physics Emulation and Hybrid Models C ☺	550
17.1.1 Radiation in Atmospheric Models C ☺	550
17.1.2 Clouds C ☺	551
17.1.3 Turbulent Fluxes C ☺	553
17.1.4 Hybrid Coupled Atmosphere–Ocean Modelling C ☺	554
17.1.5 Wind Wave Modelling C ☺	555
17.2 Physics-Informed Machine Learning A ☺	556
17.2.1 Soft Constraint C ☺	556

<i>Contents</i>	xiii
17.2.2 Hard Constraint \mathbb{C}^{\ominus}	558
17.3 Data Assimilation and ML \mathbb{C}^{\ominus}	560
17.3.1 3D-Var \mathbb{C}^{\ominus}	562
17.3.2 4D-Var \mathbb{C}^{\ominus}	563
17.3.3 Neural Networks in 4D-Var \mathbb{C}^{\ominus}	564
Exercises	568
Appendices	569
A Trends in Terminology	569
B Lagrange Multipliers	569
References	573
Index	613