Contents

Preface xi
Preface to the English edition . . . xi
A letter from the doctoral student Jiří Pecha prior to publication of the lectures . . . xii
A letter from the authors to the doctoral student Jiří Pecha . . . xiii
Basic concepts and notations . . . xv

Acknowledgements xix

Lecture 1 Bayesian statistical decision making 1
1.1 Introduction to the analysis of the Bayesian task . . . 1
1.2 Formulation of the Bayesian task . . . 1
1.3 Two properties of Bayesian strategies . . . 3
1.4 Two particular cases of the Bayesian task . . . 7
1.4.1 Probability of the wrong estimate of the state . . . 7
1.4.2 Bayesian strategy with possible rejection . . . 9
1.5 Discussion . . . 11
1.6 Bibliographical notes . . . 22

Lecture 2 Non-Bayesian statistical decision making 25
2.1 Severe restrictions of the Bayesian approach . . . 25
2.1.1 Penalty function . . . 25
2.1.2 *A priori* probability of situations . . . 26
2.1.3 Conditional probabilities of observations . . . 27
2.2 Formulation of the known and new non-Bayesian tasks . . . 28
2.2.1 Neyman–Pearson task . . . 28
2.2.2 Generalised task with two dangerous states . . . 31
2.2.3 Minimax task . . . 31
2.2.4 Wald task . . . 32
2.2.5 Statistical decision tasks with non-random interventions . 33

2.3 The pair of dual linear programming tasks, properties and solutions . . . 35
2.4 The solution of non-Bayesian tasks using duality theorems . . . 40
2.4.1 Solution of the Neyman–Pearson task . . . 41
2.4.2 Solution of generalised Neyman–Pearson task with two dangerous states . . . 44
2.4.3 Solution of the minimax task . . . 46
2.4.4 Solution of Wald task for the two states case . . . 48
2.4.5 Solution of Wald task in the case of more states . . . 50
2.4.6 Testing of complex random hypotheses . . . 52
2.4.7 Testing of complex non-random hypotheses . . . 53
2.5 Comments on non-Bayesian tasks . . . 53
2.6 Discussion . . . 54
2.7 Bibliographical notes . . . 71

Lecture 3 Two statistical models of the recognised object **73**

3.1 Conditional independence of features . . . 73
3.2 Gaussian probability distribution . . . 75
3.3 Discussion . . . 77
3.4 Bibliographical notes . . . 99

Lecture 4 Learning in pattern recognition **101**

4.1 Myths about learning in pattern recognition . . . 101
4.2 Three formulations of learning tasks in pattern recognition . . . 102
4.2.1 Learning according to the maximal likelihood . . . 104
4.2.2 Learning according to a non-random training set . . . 105
4.2.3 Learning by minimisation of empirical risk . . . 106
4.3 Basic concepts and questions of the statistical theory of learning 108
4.3.1 Informal description of learning in pattern recognition . . 108
4.3.2 Foundations of the statistical learning theory according to Chervonenkis and Vapnik . . . 113
4.4 Critical view of the statistical learning theory . . . 120
4.5 Outlines of deterministic learning . . . 122
4.6 Discussion . . . 127
4.7 Bibliographical notes . . . 136

Lecture 5 Linear discriminant function **137**

5.1 Introductory notes on linear decomposition . . . 137
5.2 Guide through the topic of the lecture . . . 138
5.3 Anderson tasks . . . 141
5.3.1 Equivalent formulation of generalised Anderson task . . . 141
5.3.2 Informal analysis of generalised Anderson task . . . 142
5.3.3 Definition of auxiliary concepts for Anderson tasks . . . 145
5.3.4 Solution of Anderson original task . . . 147
5.3.5 Formal analysis of generalised Anderson task . . . 150

5.3.6 Outline of a procedure for solving generalised Anderson task . . . 157
5.4 Linear separation of finite sets of points . . . 159
5.4.1 Formulation of tasks and their analysis . . . 159
5.4.2 Algorithms for linear separation of finite sets of points . . 163
5.4.3 Algorithm for ε-optimal separation of finite sets of points by means of the hyperplane . . . 167
5.4.4 Construction of Fisher classifiers by modifying Kozinec and perceptron algorithms . . . 169
5.4.5 Further modification of Kozinec algorithms . . . 171
5.5 Solution of the generalised Anderson task . . . 175
5.5.1 ε-solution of Anderson task . . . 175
5.5.2 Linear separation of infinite sets of points . . . 179
5.6 Discussion . . . 182
5.7 Link to a toolbox . . . 213
5.8 Bibliographical notes . . . 213

Lecture 6 Unsupervised learning **215**

6.1 Introductory comments on the specific structure of the lecture . 215
6.2 Preliminary and informal definition of unsupervised learning . . . 217
6.3 Unsupervised learning in a perceptron . . . 219
6.4 Empirical Bayesian approach after H. Robbins . . . 226
6.5 Quadratic clustering and formulation of a general clustering task 232
6.6 Unsupervised learning algorithms and their analysis . . . 238
6.6.1 Formulation of a recognition task . . . 238
6.6.2 Formulation of a learning task . . . 238
6.6.3 Formulation of an unsupervised learning task . . . 240
6.6.4 Unsupervised learning algorithm . . . 241
6.6.5 Analysis of the unsupervised learning algorithm . . . 242
6.6.6 Algorithm solving Robbins task and its analysis . . . 251
6.7 Discussion . . . 253
6.8 Link to a toolbox . . . 273
6.9 Bibliographical notes . . . 274

Lecture 7 Mutual relationship of statistical and structural recognition **275**

7.1 Statistical recognition and its application areas . . . 275
7.2 Why is structural recognition necessary for image recognition? . 277
7.2.1 Set of observations . . . 277
7.2.2 Set of hidden parameter values for an image . . . 280
7.2.3 The role of learning and unsupervised learning in image recognition . . . 281
7.3 Main concepts necessary for structural analysis . . . 284
7.4 Discussion . . . 288
7.5 Bibliographical notes . . . 305

Lecture 8 Recognition of Markovian sequences **307**

8.1 Introductory notes on sequences 307
8.2 Markovian statistical model of a recognised object 308
8.3 Recognition of the stochastic automaton 312
8.3.1 Recognition of the stochastic automaton; problem formulation 312
8.3.2 Algorithm for a stochastic automaton recognition 313
8.3.3 Matrix representation of the calculation procedure 314
8.3.4 Statistical interpretation of matrix multiplication 316
8.3.5 Recognition of the Markovian object from incomplete data 318
8.4 The most probable sequence of hidden parameters 321
8.4.1 Difference between recognition of an object as a whole and recognition of parts that form the object 321
8.4.2 Formulation of a task seeking the most probable sequence of states 321
8.4.3 Representation of a task as seeking the shortest path in a graph 321
8.4.4 Seeking the shortest path in a graph describing the task . 323
8.4.5 On the necessity of formal task analysis 326
8.4.6 Generalised matrix multiplications 327
8.4.7 Seeking the most probable subsequence of states 330
8.5 Seeking sequences composed of the most probable hidden parameters 333
8.6 Markovian objects with acyclic structure 338
8.6.1 Statistical model of an object 338
8.6.2 Calculating the probability of an observation 339
8.6.3 The most probable ensemble of hidden parameters 343
8.7 Formulation of supervised and unsupervised learning tasks 344
8.7.1 The maximum likelihood estimation of a model during learning 345
8.7.2 Minimax estimate of the model 345
8.7.3 Tuning of the recognition algorithm 346
8.7.4 Task of unsupervised learning 347
8.8 Maximum likelihood estimate of the model 347
8.9 Minimax estimate of a statistical model 353
8.9.1 Formulation of an algorithm and its properties 353
8.9.2 Analysis of a minimax estimate 356
8.9.3 Proof of the minimax estimate algorithm of a Markovian mode 366
8.10 Tuning the algorithm that recognises sequences 366
8.11 The maximum likelihood estimate of statistical model 368
8.12 Discussion 372
8.13 Link to a toolbox 395
8.14 Bibliographical notes 395

Lecture 9 Regular languages and corresponding pattern recognition tasks **397**

9.1 Regular languages . . . 397
9.2 Other ways to express regular languages . . . 399
9.2.1 Regular languages and automata . . . 399
9.2.2 Regular languages and grammars . . . 400
9.2.3 Regular languages and regular expressions . . . 401
9.2.4 Example of a regular language expressed in different ways 402
9.3 Regular languages respecting faults; best and exact matching . . 404
9.3.1 Fuzzy automata and languages . . . 405
9.3.2 Penalised automata and corresponding languages . . . 406
9.3.3 Simple best matching problem . . . 407
9.4 Partial conclusion after one part of the lecture . . . 409
9.5 Levenstein approximation of a sentence . . . 410
9.5.1 Preliminary formulation of the task . . . 410
9.5.2 Levenstein dissimilarity . . . 411
9.5.3 Known algorithm calculating Levenstein dissimilarity . . . 412
9.5.4 Modified definition of Levenstein dissimilarity and its properties . . . 414
9.5.5 Formulation of the problem and comments to it . . . 417
9.5.6 Formulation of main results and comments to them . . . 418
9.5.7 Generalised convolutions and their properties . . . 420
9.5.8 Formulation of a task and main results in convolution form . . . 427
9.5.9 Proof of the main result of this lecture . . . 429
9.5.10 Nonconvolution interpretation of the main result . . . 440
9.6 Discussion . . . 443
9.7 Link to a toolbox . . . 477
9.8 Bibliographical notes . . . 477

Lecture 10 Context-free languages, their 2-D generalisation, related tasks **479**

10.1 Introductory notes . . . 479
10.2 Informal explanation of two-dimensional grammars and languages 480
10.3 Two-dimensional context-free grammars and languages . . . 484
10.4 Exact matching problem. Generalised algorithm of C–Y–K . . . 486
10.5 General structural construction . . . 489
10.5.1 Structural construction defining observed sets . . . 490
10.5.2 Basic problem in structural recognition of images . . . 493
10.5.3 Computational procedure for solving the basic problem . 494
10.6 Discussion . . . 498
10.7 Bibliographical notes . . . 505

Bibliography **507**
Index **514**