Contents

1 Fundamentals **5**
- 1.1 Introduction . . . 5
  - 1.1.1 Organisational issues . . . 5
  - 1.1.2 Scope of this course . . . 5
  - 1.1.3 What is data analysis? . . . 6
  - 1.1.4 Role of data analysis in science . . . 7
  - 1.1.5 Role of data analysis in every-day life . . . 8
  - 1.1.6 Literature . . . 8
- 1.2 Probability . . . 9
  - 1.2.1 Definition . . . 9
  - 1.2.2 Probability mass function . . . 9
  - 1.2.3 Probability density function . . . 9
  - 1.2.4 Notation issues . . . 10
  - 1.2.5 Cumulative distribution function and characteristic function . 10
  - 1.2.6 Interpretation of probabilities . . . 10
- 1.3 Conditional probability and Bayes' theorem . . . 11
  - 1.3.1 Conditional probability . . . 11
  - 1.3.2 Marginalisation . . . 11
  - 1.3.3 Bayes' theorem and Bayesian statistics . . . 12
  - 1.3.4 Bayesian evidence . . . 13
  - 1.3.5 Bayesians vs. Frequentists . . . 14
  - 1.3.6 Choice of prior distributions . . . 15
- 1.4 Curse of dimensionality . . . 16
- 1.5 Fitting a straight line . . . 17
  - 1.5.1 Setting the stage . . . 17
  - 1.5.2 Constructing the likelihood function . . . 17
  - 1.5.3 Likelihood principle . . . 18
  - 1.5.4 Likelihood maximisation . . . 19
  - 1.5.5 Posterior maximisation . . . 20
  - 1.5.6 Real data . . . 20
- 1.6 Outlook . . . 22

2 Regression I **23**
- 2.1 Definition of "regression" . . . 23
- 2.2 Simple examples . . . 23
  - 2.2.1 Estimating a mean with Gaussian errors . . . 23
  - 2.2.2 Estimating a mean with Poisson errors . . . 25
  - 2.2.3 Estimating a fraction . . . 26
- 2.3 Central Limit Theorem . . . 26
  - 2.3.1 Taylor expansion of log-posterior . . . 26

2.3.2 Application: Fisher matrix . . . 27
2.4 Generalised linear models . . . 28
2.4.1 Linear models . . . 28
2.4.2 Exponential-family distributions . . . 28
2.4.3 Generalisation of linear models . . . 30
2.4.4 Global likelihood maximum . . . 30
2.4.5 Salpeter stellar mass function . . . 31
2.5 Summary: Where are we? . . . 33

3 Regression II **35**
3.1 Nonlinear models . . . 35
3.2 Regression under constraints . . . 36
3.2.1 Constraints vs. priors . . . 36
3.2.2 Salpeter stellar mass function revisited . . . 37
3.2.3 Linear regression on a plane . . . 38
3.2.4 Regression under positivity constraints . . . 40
3.3 Limitation of classical regression . . . 41

4 Numerical methods **43**
4.1 Monte-Carlo sampling . . . 43
4.1.1 Numerical integration . . . 43
4.1.2 Sampling from the cumulative distribution . . . 44
4.1.3 Built-in samplers . . . 45
4.1.4 Rejection sampling . . . 46
4.1.5 Importance sampling for integration . . . 47
4.2 Simple optimisation algorithms . . . 48
4.2.1 Brute-force grids . . . 48
4.2.2 Gradient descent . . . 48
4.2.3 Newton's method . . . 49
4.2.4 Schechter galaxy luminosity function . . . 50
4.3 Advanced optimisation algorithms . . . 51
4.3.1 Metropolis-Hastings . . . 51
4.3.2 EMCEE . . . 53
4.3.3 Other MCMC algorithms . . . 54
4.4 Efficient optimisation algorithms . . . 55
4.4.1 Big Data . . . 55
4.4.2 Stochastic gradient ascent . . . 55
4.4.3 Stochastic EMCEE . . . 57
4.5 Summary: Where are we? . . . 57

5 Model comparison **59**
5.1 Underfitting and overfitting . . . 59
5.1.1 Generalisation error . . . 59
5.1.2 Under- and overfitting in parameter estimation . . . 60
5.1.3 Under- and overfitting and model complexity . . . 61
5.1.4 How to handle overfitting . . . 62
5.1.5 Bias-variance decomposition . . . 62
5.2 Cross validation . . . 63
5.3 Bayesian model comparison . . . 64
5.3.1 Bayesian evidence and goodness-of-fit . . . 64
5.3.2 Bayes factors . . . 64

5.3.3 Analytic evidence for linear models . . . 65
5.3.4 Connection between AGN and galaxy merging . . . 67
5.3.5 Estimating evidence numerically . . . 68
5.3.6 Robust and efficient evidence estimation from MCMC . . . 70
5.4 The failure of Frequentist model assessment . . . 72
5.4.1 Orthodox hypothesis testing . . . 72
5.4.2 Reduced χ^2 . . . 73
5.4.3 χ^2-test as an example of p-values . . . 74
5.4.4 Problems of p-values . . . 75
5.4.5 Demonstration with toy data . . . 76
5.4.6 Discussion with regard to Gliese 581 . . . 76
5.5 Predicting data . . . 77
5.5.1 Single-model prediction . . . 77
5.5.2 Multi-model prediction . . . 78
5.6 Summary: Where are we? . . . 79

6 Classification **81**
6.1 About classification . . . 81
6.1.1 Classification in astronomy . . . 81
6.1.2 Supervised vs. unsupervised learning . . . 81
6.2 Perceptron classifier and logistic regression . . . 82
6.2.1 Motivation . . . 82
6.2.2 The perceptron classifier . . . 82
6.2.3 The logistic regression classifier . . . 83
6.2.4 Gradient and Hessian of logistic regression . . . 83
6.3 Neural networks . . . 84
6.3.1 Perceptron networks . . . 84
6.3.2 Building neural networks . . . 86
6.3.3 Predictions from neural networks . . . 87
6.3.4 Training neural networks . . . 88
6.3.5 Problems of neural networks . . . 89
6.4 Other modern classification algorithms . . . 90
6.4.1 Nearest-neighbour in high dimensions . . . 91
6.5 Unsupervised classification . . . 92
6.5.1 K-means clustering . . . 92
6.5.2 Other unsupervised learning algorithms . . . 93
6.5.3 Practical problems . . . 93
6.6 Summary: What have we learned? . . . 94

7 Learning Theory **95**
7.1 What Learning Theory is about . . . 95
7.2 The learning problem . . . 96
7.2.1 Ingredients . . . 96
7.2.2 Risk functional and its special cases . . . 96
7.2.3 Empirical risk minimisation . . . 97
7.2.4 Questions for Learning Theory . . . 98
7.3 Learning Theory in a nutshell . . . 98
7.3.1 Milestones of Learning Theory . . . 98
7.3.2 VC dimension . . . 98
7.3.3 Fast convergence . . . 99

7.3.4 Evaluating the VC dimension 100
7.4 Structural risk minimisation 102
7.4.1 Training error and generalisation 102
7.4.2 Sets of functions with structure 103
7.4.3 Maximum-margin classifier 103
7.4.4 Support Vector Machine 105