Intelligent Data Analysis
Exam questions

Course text book book: Bishop, C.M., 2006. *Pattern recognition and machine learning*. Springer. Publicly available at:
https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

1 **-** lecture
2 **-** further reading belonging to the lecture
3 **-** background reading to improve on basics

I. Phases of data analysis using machine learning, Visual Exploratory Data Analysis

Knowledge acquisition, study design, data collection, quality control, outlier/anomaly detection, data imputation, data engineering (dimensionality reduction, feature selection), method and (hyper)parameter selection, integration of the results (fusion), decision systems, IT embedding, deployment, application. Goals and main plots & how to use them for preparing deep analysis questions.

https://en.wikipedia.org/wiki/Exploratory_data_analysis
https://www.interkonyv.hu/konyvek/antal_peter_intelligens_adatelemzes : 5. fejezet Kocsis Imre:Vizuális analízis
Bishop: 1 Introduction (2)

II. Paradigms of learning

Inductive inference (induction) by Bayesian model averaging (concept of prior, likelihood, posterior). (Full) Bayesian learning and Maximum A Posteriori (MAP)/ Maximum Likelihood (ML) principles, models. Example with finite models and binary data.

https://en.wikipedia.org/wiki/Bayesian_inference
https://en.wikipedia.org/wiki/Maximum_likelihood_estimation
Bishop: 1.1 Example: Polynomial Curve Fitting (1)
14 Combining Models
14.1 Bayesian Model Averaging (1)

III. Decision theory

Optimal decision. Classification. Formulation of the problem. Basic concepts: false/true x positive/negative decisions; sensitivity, specificity, positive/negative predictive value, Receiver Operating Characteristics Curve (ROC curve), Area Under the ROC curve (AUC).

https://en.wikipedia.org/wiki/Statistical_classification
https://en.wikipedia.org/wiki/Bayes_error_rate
https://en.wikipedia.org/wiki/Optimal_decision
https://en.wikipedia.org/wiki/Receiver_operating_characteristic
Bishop: 1.5 Decision Theory (1)
1.5.1 Minimizing the misclassification rate (1)
1.5.2 Minimizing the expected loss (1)
1.5.3 The reject option (1)
1.5.4 Inference and decision (1)

IV. Linear models for regression, Maximum likelihood and least squares

Linear regression. Linear-in-parameters models. The basic principle of LS, ML, and MAP (Bayesian) approaches. The LS solution. Problem of overfitting

https://en.wikipedia.org/wiki/Linear_regression
Bishop: 3 Linear Models for Regression (1)
3.1 Linear Basis Function Models (1)
3.1.1 Maximum likelihood and least squares (1)

V. Regularization in learning, Bias/Variance trade-off

The role of regularization. L2 and L1 regularizations. Regularized LS solution. Notion of bias, variance and mean square error. Trade-off and regularization.

https://en.wikipedia.org/wiki/Regularization_(mathematics)
https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff
Bishop: 3.1.4 Regularized least squares (1)
3.2 The Bias-Variance Decomposition (1)

VI. *Linear models for classification*

Linearly separable problems and decision boundaries.

Bishop: 4 Linear Models for Classification (1)
4.1 Discriminant Functions (1)
4.1.1 Two classes (1)
4.1.2 Multiple classes (1)
4.1.3 Least squares for classification (2)
4.1.4 Fisher's linear discriminant (1)

VII. *Classification with probabilistic generative and discriminative models*

Bishop: 4.2 Probabilistic Generative Models (1)
4.2.1 Continuous inputs (1)
4.2.2 Maximum likelihood solution (1)
4.3 Probabilistic Discriminative Models (1)
4.3.1 Fixed basis functions (2)
4.3.2 Logistic regression (1)

VIII. Clustering and Expectation-Maximization (E-M)

K-means. Hierarchical clustering.
https://en.wikipedia.org/wiki/K-means_clustering
https://en.wikipedia.org/wiki/Hierarchical_clustering
Bishop: 9 Mixture Models and EM (1)
9.1 K-means Clustering (1)
9.2 Mixtures of Gaussians (1)
9.2.1 Maximum likelihood (1)
9.2.2 EM for Gaussian mixtures (1)

IX. *Neural networks*

https://en.wikipedia.org/wiki/Artificial_neural_network
https://en.wikipedia.org/wiki/Multilayer_perceptron
Bishop: 5 Neural Networks (1)
5.1 Feed-forward Network Functions (1)
5.2 Network Training (1)
5.2.1 Parameter optimization (1)
5.2.2 Local quadratic approximation (1)
5.2.3 Use of gradient information (1)
5.2.4 Gradient descent optimization (1)
5.3 Error Backpropagation (1)
5.3.1 Evaluation of error-function derivatives (1)
5.3.2 A simple example (1)
5.5 Regularization in Neural Networks (1)

X. Independence models of probability distributions.

Concept of independence and properties of independencies. The Markov Blanket set. Independence assumptions of naive Bayesian networks and Hidden Markov Models
https://en.wikipedia.org/wiki/Conditional_independence
https://en.wikipedia.org/wiki/Naive_Bayes_classifier
https://en.wikipedia.org/wiki/Hidden_Markov_model
https://en.wikipedia.org/wiki/Markov_blanket
Bishop: 8.2 Conditional Independence
8.2.1 Three example graphs
8.2.2 D-separation

XI. Bayesian networks

Factorization based construction of a Bayesian network. Concept of d-separation. Observational equivalence classes of Bayesian networks
https://en.wikipedia.org/wiki/Bayesian_network#d-separation
https://en.wikipedia.org/wiki/Bayesian_network#Structure_learning
Bishop: 8.2 Conditional Independence
8.2.1 Three example graphs

8.2.2 D-separation
8.4.8 Learning the graph structure

XII. Markov Chain Monte Carlo methods

Markov Chain Monte Carlo (MCMC) methods (in finite state space). Irreducibility, aperiodicity, detailed balance. The Metropolis-Hastings sampling.

https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo

Bishop: 11.2 Markov Chain Monte Carlo
11.2.1 Markov chains
11.2.2 The Metropolis-Hastings algorithm
11.3 Gibbs Sampling