Παράρτημα Ι

Προπαρασκευαστικό μάθημα
Data programming *Topics per week* 1. Introductory programming concepts 2. Python basics and built-in data structures 3. Object-oriented programming basics 4. Data lifecycle and properties of good software 5. Algorithms
Α′ εξάμηνο, υποχρεωτικά μαθήματα:
Data management *Topics per week* 1. Overview 2. Entity relation model 3. Relational model 4. Relational algebra 5. SQL 6. Query processing 7-8. Query optimisation 9. Primary and secondary storage 10. Tree-structured indexes 11. Hash-based structures 12-13. Database tuning and physical design for massive datasets
Large-scale statistical methods *Topics per week* 1. Review on basic probability theorems 2. Discrete and continuous random variables 3. Bayesian inference and the posterior distribution 4. Point estimation, hypothesis testing, and the MAP Rule 5. Bayesian least mean squares estimation 6. Bayesian linear least mean squares estimation 7. Statistical inference 8. Classical parameter estimation 9. Linear regression 10. Binary hypothesis testing 11. Significance testing 12-13. Introduction to multivariate models
Machine learning *Topics per week* 1. Introduction 2. Regression 3. Decision Trees 4. Logistic Regression 5. Part 1: kNN & clustering Part 2: Evaluation & Performance measures

6. Applied Machine Learning I
7. Applied Machine Learning II
8. Support Vector Machines
9. Naive Bayes
10. Ensembles
11. Feature Selection and Dimensionality Reduction
12. Dimensionality Reduction and Gradient Descent
13. Sampling

Big data mining

Topics per week

1. Data mining basic concepts
2. Data types and features
3-4. Data preprocessing and cleaning
5-6. Data classification and clustering
7. Itemset mining
8. Outliers and concept drift
9. Evaluation in data mining
10. Introduction to Natural Language Processing
11. Morphology and language models
12. Vectors semantics and neural representations
13. Syntactic and semantic parsing

Β′ εξάμηνο, υποχρεωτικά μαθήματα:

Big data management

Topics per week

1. Getting to know your (Big) Data
2-3. Architectures for Big Data
4. Distributed object location
5. Distributed file systems (Cassandra, BigTable, HBase)
6. The Map/Reduce paradigm
7-9. Parallel data processing with Hadoop
10. Parallel graph processing (Pregel, Hama)
11. NoSQL databases (key-value/document/graph stores)
12. Column stores
13. Distributed stream processing

Big data security and visualization

Topics per week

Big data security

1. Introduction to security
2. Cryptography for big data
3. Distributed systems security
4. Dynamic risk models
5. Large network security
6. Intrusion detection systems
7. Distributed trust

Data visualization

8. Visual perception

9-10. Visualization techniques
11. Interaction techniques
12-13. Visualization software (Tableau, Python)

Applied data science

Topics per week

1. Scientific method overview
2. Hypotheses and testing
3. Risks in hypothesis testing
4. Scientific error and scientific lies
5. Reviewing scientific work: the peer reviewing process; how to do a good review; how to review one's own work.
6. Communicating scientific results: clarifying science; risks in publication of results
7. Legal and ethical issues overview: overview of legal and ethical risks
8. Data licensing, sharing, openness: how to share or reuse data; licences and their meaning
9. Emerging data formats and publishing (nano-publications; semantic web)
10. Anonymization and profiling: data aggregation and anonymization; discovering user identity through profiling
11. Privacy and Security concerns: difference between privacy and security; privacy in data publication; sensitive data
12. Ethics considerations in data analysis: the effect and impact of scientific discovery; ethics and data analysis
13. Social understanding of data and ethics

Deep learning

Topics per week

1. Introduction
2. Feedforward neural nets, backprop, regularisation
3. Optimisation and practical issues
4. Convolutional networks
5. Recurrent and Recursive Networks
6. Autoencoders and representation learning
7. The Long Short-Term Memory and Other Gated RNNs
8. Using external categorical evidence for clustering
9. Sequence Models and Attention
10. Laboratory
11 Inductive Transfer
12. Data Augmentation
13. Visiting lecture: CNN Architectures for Object Detection