Data Science MSc (2023) – final exam topics

Fundamentals of the theoretical background of data science:

1. Basic concepts of information security (CIA triad); Means and objectives of cyber security; Malware, attack technologies; Access control (DAC, MAC, RBAC, ABAC, CBAC); Access control in distributed systems.
2. Authentication, user authentication; Authentication in distributed systems; Accountability; Security operations and incident management; Monitor, Analyze, Plan, Execute and Knowledge (MAPE-K); Regulations, standards.
3. Supervised and unsupervised learning; Univariate and multivariate linear regression; Gradient descent method; Stochastic and mini-batch gradient descent; Feature normalization; Polynomial regression; Normal equation; Logistic regression; Binary and multiclass classification; Regularization (underfitting and overfitting); Regularized linear and logistic regression; Neural networks; Backpropagation algorithm; Numerical gradient checking.
4. Training/Test/Validation data split; Training diagnostics; Learning curves (training set size); Error measurement and skewed classes; Application of support vector machines and kernel functions; Clustering; Determining the number of clusters; Dimensionality reduction; Anomaly detection (normal distribution); Recommender systems; Content-based recommendation; Collaborative filtering; Map-Reduce and parallelization.
5. Unconstrained and constrained extrema of multivariate functions; Gradient methods; Trust region; Newton's method, quasi-Newton methods; Conjugate gradient method; Least squares method; Stochastic optimization.
6. Multivariate sample and its properties, multivariate normal distribution; Principal component analysis; Factor analysis; Canonical correlation analysis; Classification methods (maximum likelihood and Bayesian classification, linear and quadratic discriminant analysis, nearest neighbor method); Multidimensional scaling.

Fundamentals of the practical background of data science:

1. Description of the cloud infrastructure; Distinguishing between IaaS, PaaS, and SaaS; Differences between the types of cloud computing (public, private, on-premises, hybrid); Reliability, availability, and scalability of cloud environments; Analysis of the costs of cloud applications and systems; Availability metrics in cloud environments.
2. Overview of the differences between public, private and hybrid cloud application deployments; Creating a virtual environment (virtual machines, virtual network, data storage); Managing operational costs; Identifying potential risks and disaster scenarios; Developing on-premises and off-premises backup strategies; Monitoring cloud-based systems; Managing cloud security.
3. Basic concepts of data visualization and the history of its development, the role of visual perception; Data abstraction, data types, data preparation for visualization; Tasks and goals, task abstraction; The design process of visualization; Visualizing categorical, numerical, ordinal, time series, and geographical data; Visualization of trees, graphs, and networks; Interaction, scalability, animation, colors.
4. Big data visualization techniques; Dimensionality reduction techniques; Importance of visualization in exploratory data analysis; Creating dashboards; Storytelling; Current libraries and software for data visualization.
5. Theory and evolution of data protection in the European legal culture; Pool of personal data; Data controllers and data processors; Principles of the GDPR; Rights of the data subjects; Data breach, reporting obligation, investigation, role of the national data protection authorities; Legal consequences of the infringement of personal data protection; ePrivacy in the European Union.
6. The role of codes of conduct in data protection; Freedom of information and the publicity of public interest data; CJEU case law on data protection; Relationship between data protection and ethics; Big Data: ethical and legal concerns; IoT: ethical and legal concerns.
7. Libraries related to artificial intelligence programming; Kernel-level runtime and compiler environments; Common libraries and tools (scikit-learn, NumPy, SciPy, pandas, Jupyter, Matplotlib, Dataflow, Keras, TensorFlow).