Data Science Track Final Exam Topics

Introduction to Data Science 1.

- Types of attributes
  - e.g., I./1.
- Similarities, distances (Minkowski distances, Mahalanobis distance, Simple Matching Coefficient (SMC), Jaccard coefficient, cosine distance, Dynamic Time Warping (DTW))
  - e.g., I./2., 3., 4., 5., 6., 7.
- Construction and operation of decision trees (misclassification error, Gini index, based on entropy)
  - e.g., II./2., 3., 5., III./4.
- Evaluation of classification, performance metrics (precision, recall, F-measure, ROC, AUC)
  - e.g., II./3., 4., 5.
- Naive Bayes method (Laplace estimation, m-estimation)
  - e.g., II./6.
- Demonstration of the operation and performance evaluation of classification algorithms (kNN, decision tree, Naive Bayes) on visually provided small sample data
  - e.g., II./7., 8.
- Calculations with Bayesian networks, conditional independence
  - e.g., III./1.
- Linear regression
  - e.g., III./2.
- Gradient descent method, stochastic gradient descent method
  - e.g., III./2.
- Operation of perceptron, construction of neural networks
  - e.g., III./3., 6., 8.
- Linear separability, maximum margin principle
  - e.g., III./5., 7.
- Association rules (support, confidence, lift metric, Apriori algorithm, maximal, closed itemsets)
  - e.g., IV./1.
- Operation of k-means clustering
  - e.g., IV./2., 4., 6.
- Operation of hierarchical clustering (MIN, MAX distance)
  - e.g., IV./4., 5.

- Demonstration of the operation and performance evaluation of clustering algorithms (hierarchical clustering (in case of MIN or MAX distance), k-means, Gaussian Mixture Model, DBSCAN) on visually provided small sample data
  - e.g., IV./7.
- Recommender system with latent factor model
  - e.g., IV./3.
- Operation of the PageRank algorithm
  - e.g., IV./8., 9., 10.