KDDM so far

This sheet shall give you an idea of which contents have been covered by the lecture and will also provide a **most likely incomplete** list of relevant contents to support (or somewhat guide) your review of the lecture and exercise materials.

1

- Data Mining:
    - term
    - disciplines involved
    - data mining vs machine learning
    - motivations for data mining
- Knowledge Discovery in Databases
    - term and core aspects
    - KDD process and steps
    - supervised vs. unsupervised learning
    - tasks: clustering, classification, regression, association rule mining, outlier detection

2

- data preprocessing and feature spaces
    - steps and tasks of preprocessing and transformation
- data sets
    - instances
    - features
- basic feature types: value range, applicable relations, differences
    - binary
    - categorical (nominal, ordinal)
    - numerical (interval-scaled, ratio-scaled)
- data descriptors and visualization
    - univariate: mean, median, mode, skewness, variance, standard deviation, percentiles
    - bivariate: correlation coefficient, contingency table, chi-square
- feature spaces and proximity
    - feature space, metric space (formal difference)
    - similarity vs distance measure
    - proximity measures (L-norms)
    - normalization: why? how?
    - text data: challenges and approaches

3

- concepts in frequent itemset mining
  - item, itemset, itemset size, k-itemset, transaction, database, lexicographical order, itemset lattice
  - cover, absolute support / support count, relative support, frequent itemset, L
  - association rule, support of a rule, confidence
- problem settings
  - FIM, ARM
- FIM: Apriori
  - Approach and algorithm, apriori property, pruning
- Association Rule Mining
  - approach, interest, lift, confidence

4

- Apriori improvements
  - apriori challenges, improvements
- FP-Growth
  - Approach and algorithm
  - advantages of FP-Growth
- DB scanning costs
  - Partition
  - Sampling
  - Eclat
- too many frequent itemsets
  - closed frequent itemsets (CFI)
  - maximal frequent itemsets (MFI)