TU Dortmund 8.7.2022
Artificial Intelligence / Data Science and Data Engineering **Machine Learning for Complex Data**
Prof. Dr. Erich Schubert
Prof. Dr. Emmanuel Müller **Summer 2022**
Daniel Wilmes

Assignment 11: EDSC and RESCU

Due: Thursday, 14.7.2022

Problem 11-1 EDSC: Hypercube monotonictiy 3

Prove the following Theorem:
For any two subspaces with $\mathbf{S} \subseteq \mathbf{T}$, and any hypercube $\mathbf{H}$ identified by interval indices $[a_1,b_1]...[a_d,b_d]$ the number of objects in $\mathbf{T}$ is bound by the number of objects in $\mathbf{S}$:

$$|\mathbf{H}^{\mathbf{S}}_{[\mathbf{a_1},\mathbf{b_1}]...[\mathbf{a_d},\mathbf{b_d}]}| \geq |\mathbf{H}^{\mathbf{T}}_{[\mathbf{a_1},\mathbf{b_1}]...[\mathbf{a_d},\mathbf{b_d}]}|$$

Problem 11-2 EDSC: Completeness of density filter 3

Prove the following Theorem: For all subspaces $\mathbf{S}$ with $\mathbf{S} \subseteq \mathbf{T} \subseteq \mathbf{D}$ and $|\mathbf{T}| = t$ holds:

If an object is not **weak dense** in $\mathbf{S}$, it is not **dense** in $\mathbf{T}$.

Problem 11-3 Set covers and RESCU 8

(a) The *set cover problem* can be stated as follows:
Given the set $U = \{1,...n\}$ (called the *universe*) and the set $S = \{S_1,...,S_m\}$ where $S_i \subseteq U$ for all $S_i \in S$, find a minimal collection of these subsets whose union is U (i.e. a set cover of minimum size).

Given the following greedy algorithm:
Select the largest subset of S, and then delete all its elements from U. Repeat by adding the subset containing the largest number of uncovered elements until all elements of U are covered.

Construct an example where the greedy solution yields a different solution then the optimal solution. State the optimal and the greedy solution and all intermediate steps.

(b) The *weighted set cover problem* can be stated as follows:
Given the set $U = \{1,...n\}$ (the *universe*) and the set $S = \{S_1,...,S_m\}$ where $S_i \subseteq U$ for all $S_i \in S$, each S_i is assigned a nonnegative cost $cost(S_i)$. Find a minimum-cost collection of these subsets whose union is U (i.e. a set cover of minimum cost).

Given the following greedy algorithm:
Select the subset of S which has the lowest cost, and then delete all its elements from U. Repeat by adding the lowest-cost-subset which contains uncovered elements until all elements of U are covered.

Construct an example where the greedy solution yields a different solution then the optimal solution. State the optimal and the greedy solution and all intermediate steps. (Use $cost(S_i) = 1/|S_i|$ as cost function.)

(c) How do the set cover problems relate to the RESCU algorithm?