

ADBSÜ Gruppe 3

Sampling Algorithmen

Sampling

- Repräsentatives Beispiel eines Datensatzes halten
- Repräsentativ: Uniform zufällig
- Probleme
 - Löschen
 - **Größe des Samples**
 - Zeitfenster
 - Verteilte Datensätze

Grundlegende Algorithmen

- Reservoir Sampling
 - N Samples
 - 1 Durchlauf
 - Kein DELETE
- Priority Sampling
 - Neuer Datensatz bekommt zufällige Priorität
 - Nicht ausgelaufener DS mit höchster Priorität ist Sample
 - Mehrere Samples durch mehrere Sampler

Fortgeschrittene Algorithmen

- Random Pairing
 - Stabile Samplegröße
 - Stabile Datensatzgröße
 - Kein Zugriff auf Datensatz nötig
- Bounded-Space Priority Sampling w/o Replacement
 - Menge an k Kandidaten und Test-Items
 - Sample $S = \text{topk}(S_{cand} \cup S_{test}) \cap S_{cand}$

Demo

To do

- Random Pairing
 - **Bugfixes**, Tests
 - GUI
 - Optimized RP implementieren
- BPSWOR
 - **Bugfixes**, Tests
 - Templates