

## Aufgabenblatt 1 (Deskriptive Methoden)

### Aufgabe 1 – Eigenschaften von Attributen

(20 Punkte)

Gegeben seien die folgenden Werte und ihre absolute Häufigkeit

| Wert | Absolute Häufigkeit |
|------|---------------------|
| 1    | 2                   |
| 2    | 4                   |
| 3    | 1                   |
| 4    | 1                   |

- (a) Bestimmen Sie Varianz und Standardabweichung. (4 Punkte)
- (b) Bestimmen Sie die Schiefe. (4 Punkte)
- (c) Bestimmen Sie die empirische Wölbung. (4 Punkte)
- (d) Standardisieren Sie die obigen Werte. (4 Punkte)
- (e) Min-Max-skalieren Sie die obigen Werte. (4 Punkte)

### Aufgabe 2 – Verständnisfragen

(42 Punkte)

- (a) Visualisieren Sie ein Parallelkoordinaten-Diagramm für 2 korrelierte Attribute mit 4 beliebigen Beobachtungen. (2 Punkte)
- (b) Erläutern Sie anhand der Datenerzeugung, Datenverarbeitung und Datennutzung wie Big Data Probleme entstehen und grenzen Sie dies von traditionellem Data Mining ab. (6 Punkte)
- (c) Nennen und beschreiben Sie kurz typische Fehler in Daten. (5 Punkte)
- (d) Nennen Sie die Ihnen bekannten Skalentypen und diskutieren Sie, wie diese für Entscheidungsbäume genutzt werden können. (6 Punkte)
- (e) Beschreiben Sie kurz die Grundidee von Biased Sampling und Stratified Sampling und geben Sie je ein Szenario an, für das die Sampling Methoden besonders geeignet sind. (6 Punkte)
- (f) Beschreiben Sie Struktur und Nutzen von Taxonomien für die Datenqualität. (4 Punkte)
- (g) Nennen Sie zwei andere Klassifikationsverfahren neben den Entscheidungsbäumen und diskutieren Sie kurz die Vorteile der Verfahren gegenüber Entscheidungsbäumen. (4 Punkte)
- (h) Nennen Sie vier verschiedene Stoppkriterien für Entscheidungsbäume. (4 Punkte)