

Übungsblatt 3, Business Analytics, SoSe 2012, 06.07.2012
Dr. Tomáš Horváth, Osman Akcatepe

1. Normalisieren Sie das Attribut X durch
 - (a) Dezimalskalierung, wenn es zwischen 0.88 und 999.99 spannen würde, falls wir $j = 3$ ansetzen würden,
 - (b) eben Dezimalskalierung, wenn die Spannweite von Attribute X von -500 zu 45 mit $j=3$ ist,
 - (c) Min-Max Normalisierung zum Intervall $[0,1]$, wenn sein Wert 40 im Intervall $[20,100]$ ist, welche transformiert sein muss.
2. Sei Dataset $D = \{13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70\}$, welche das Attribut Alter enthält. Benutzen Sie Binning mit dem arithmetischen Mittel durch ein Bin mit der Tiefe von 3, um die obige Daten zu glätten. Illustrieren Sie Ihre Einzelschritte.
3. Basierend auf dem Dataset D für das Alter im Beispiel 2 antworten Sie folgende:
 - (a) Benutzen Sie Min-Max Normalisierung den Wert 35 ins Intervall $[0.0, 1.0]$ zu transformieren.
 - (b) Benutzen Sie z-score Normalisierung den Wert 35 zu transformieren, so dass die Standardabweichung 12.94 ist.
 - (c) Benutzen Sie Dezimalskalierung den Wert 35 für das Alter zu transformieren.
 - (d) Vergleichen Sie die obige Normalisierungsmethoden zur Interpretation, so dass welche Methode und warum Sie für die gegebene Daten bevorzugen würden.
4. Basierend auf dem Dataset D für das Alter im Beispiel 2, geben Sie ein Beispiel für die stratifizierte Zufallsstichprobe (stratified sampling) mit der Aufteilung im Verhältnis. Benutzen Sie die Zufallsstichprobe von der Massangabe 13 und die Gruppen für
 - (a) Alter < 22 und
 - (b) Alter ≥ 22 .

5. Sei $N = 10$, $\phi(i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$, $N\left(\frac{2N}{5}, \frac{N}{10}\right)$, $f(n)=n^2$. Finden Sie die minimale Kosten und berechnen Sie die Historien (schedules) entsprechend den minimalen Kosten.

Bonus: Schreiben Sie ein Programm mit $N=100, 1000, 10K, 100K$ durch die oben erwähnte Verteilung und Abbildung, um die Historie zu berechnen. Wie sieht das Resultat aus?

**Viel
Erfolg !**