

Sportinformatik

WS 2019/2020 Übung 9

Abgabetermin: Mittwoch, 8. Januar 2020, 8 Uhr

Aufgabe 1: Ranglisten im Fußball - ELO Ranking

14 Punkte

1. Schreiben Sie eine Funktion, die ein Data Frame mit den Spalten `Date`, `HomeTeam`, `AwayTeam`, `FTHG` (Full Time Home Team Goals), `FTAG` (Full Time Away Team Goals) übergeben bekommt und für alle Spiele im Zeitraum für alle Mannschaften das **ELO-Ranking** berechnet. Die Funktion soll sowohl den Punktstand am Ende des Zeitraums, als auch ein Data Frame mit den Punktenständen der Mannschaften während des gesamten Zeitraums zurückgeben. 6 Punkte
 2. Verwenden Sie die Funktion aus 1.1 um für die kompletten Datensätze von Aufgabenblatt 8 Aufgabe 1 und Aufgabe 2 das Ranking zu bestimmen. Nutzen Sie ihre Funktion aus Aufgabenblatt 8 Aufgabe 3.2, um jeweils den finalen Punktstand in eine Datei zu schreiben. 1 Punkte
 3. Erzeugen Sie mit ihrer Funktion aus Aufgabenblatt 8 Aufgabe 3.2. für die 10 besten Mannschaften der beiden Datensätze jeweils einen Plot mit dem Verlauf des Rankings. 1 Punkte
 4. Wiederholen Sie Aufgabe 1.2 und 1.3 nur für die Bundesliga Spiele der Saison 2018/2019. Kommentieren Sie ihr Ergebnis im Vergleich mit dem klassischen 3-Punkte Ranking. 1 Punkte
 - a) Testen Sie unterschiedliche Werte für k . Welcher Wert liefert das plausibelste Ergebnis? 1 Punkte
 - b) Verwenden Sie zusätzlich die Daten weiterer Saisons vor 2018/2019. Wie verändert sich das Ergebnis? 1 Punkte
 5. Konvergenz des ELO-Rankings: Betrachten Sie nur die Bundesligaspiele der Saison 2018/2019. Berechnen Sie mehrfach das ELO-Ranking. Starten Sie dabei mit dem Ergebnis der letzten Berechnung als neue ELO-Startwerte. Konvergiert das Verfahren gegen ein ELO-Ranking? Variieren Sie k . Gibt es einen Unterschied? 4 Punkte
(Hinweis: Den Verlauf der ELO-Werte benötigen Sie für diesen Aufgabenteil nicht.)
-

Aufgabe 2: Distance Metrics**14 Punkte**

Seien $\tau_A(x)$ und $\tau_B(x)$ zwei Rankings für Elemente $x \in M$.

Wie stark sich τ_A und τ_B unterscheiden kann mit den folgenden Metriken berechnet werden:

Spearman's footrule:

$$d_s(\tau_A, \tau_B) = \sum_{x \in M} |\tau_A(x) - \tau_B(x)|$$

Kendall tau distance:

$$d_k(\tau_A, \tau_B) = |\{(x, y) \in M \mid x < y, (\tau_A(x) < \tau_A(y) \text{ **und** } \tau_B(x) > \tau_B(y)) \\ \text{ **oder** } (\tau_A(x) > \tau_A(y) \text{ **und** } \tau_B(x) < \tau_B(y))\}|$$

1. Schreiben Sie eine Funktion `spearman's(R1, R2)`, die Spearman's footrule für Rankings R1 und R2 berechnet. 2 Punkte
 2. Schreiben Sie eine Funktion `kenndalls(R1, R2)`, die die Kendall tau Distanz für Rankings R1 und R2 berechnet. 3 Punkte
 3. Berechnen Sie das ELO Rating am letzten Spieltag der Saison 2018/2019 mit unterschiedlichen Werten für k . Kombinieren Sie die k 's mit unterschiedlich weit in die Vergangenheit reichenden Datensätzen (1 Saison, 2 Saisons usw.). Insgesamt sollen die ELO Ratings für mindestens 20 verschiedene Kombinationen aus Seasons und k 's berechnet werden. 3 Punkte
 4. Berechnen Sie Spearman's footrule für die unterschiedlichen ELO Ratings aus 3. verglichen mit dem "normalen" 3-Punkte Ranking für die Saison 2018/2019. Schreiben Sie die Ergebnisse in eine Tabelle und sortieren sie diese aufsteigend nach der Differenz. 2 Punkte
 - a) Welche Kombination aus k 's und Saisons liegt am nächsten am 3-Punkte Ranking? 1 Punkte
 - b) Plotten Sie für die fünf besten Mannschaften im 3-Punkte Ranking die Änderung im Ranking über den Spieltagen der letzten Saison. Fügen Sie dem Plot die Ergebnisse für dieselben Mannschaften aus dem nächsten ELO-Rating hinzu. 2 Punkte
 5. Wiederholen Sie das Verfahren aus 4. für die Kendall tau Distanz. 1 Punkte
-

Aufgabe 3: Eigenvektorranging

9 Punkte

James P. Keener beschreibt in “[The Perron-Frobenius Theorem and the Ranking of Football Teams](#)” weitere Ranking Verfahren

1. Berechnen Sie das in “2. The direct method” beschriebene Ranking für die Saison 2018/19. Nutzen Sie dafür die in (2.3) beschriebene “power method”. 6 Punkte
 2. Modifizieren Sie den Ansatz aus (2.3) indem Sie für die Matrixeinträge statt 0, 0.5 und 1 die auf S. 83 in Zeile 4 angegebene Formel $a_{ij} = (S_{ij} + 1) / (S_{ij} + S_{ji} + 2)$ verwenden. (Die Modifikation aus (2.4) ist nicht notwendig.) Wie ändert sich das Ranking? 2 Punkte
 3. Wie unterscheiden sich die Ergebnisse aus 1. und 2. im Vergleich zum 3P-Ranking und ELO-Ranking? 1 Punkte
-

Gesamtpunktzahl:

37 Punkte