

Wichtig: Antwortblatt downloaden und auf deinem PC ausfüllen!

Antwortblatt zum Arbeitsblatt 1: Datenanalyse

Aufgabe 1 | Beantworte die folgenden Fragen

1. Was ist der beliebteste Film in dem Datensatz?

2. Aus welchem Jahr stammen die ältesten Filme in dem Datensatz? Aus welchem Jahr die jüngsten?

3. Wie hat sich die Verteilung der Bewertungen im Laufe der Erscheinungsjahre geändert?

4. Welcher Film wurde am häufigsten bewertet?

5. Bei welchem Film weicht die Bewertung der Nutzer am wenigsten vom durchschnittlichen Rating ab?

Aufgabe 2 | Die Rating-Tabelle

Teil a

Sammle Ideen, wie man die **bekannten Bewertungen** aus der Bewertungstabelle nutzen kann, um einem User Filme zu empfehlen, die er noch nicht gesehen hat.

Teil b

Welche Zusammenhänge zwischen verschiedenen Nutzern oder Filmen kannst du in der Rating-Tabelle finden? Beschreibe, was dies für die jeweiligen Nutzer bzw. Filme bedeutet.

Teil c

Wie würdet ihr die fehlenden Bewertungen vorhersagen? Begründet eure Antwort

Zusatzaufgabe

Wie viel Prozent der Bewertungen im Netflixdatensatz sind bekannt? Wie viel Prozent sind nicht bekannt?

Antwortblatt zu Arbeitsblatt 2 | Modell zur Vorhersage

Aufgabe 2 | Formeln für die Berechnung eines Ratings

Teil b | Beliebige Eigenschaften

Formel für den Eintrag r_{34}

$$R = \begin{matrix} & \begin{matrix} \text{Filmsymbol}_1 & \text{Filmsymbol}_2 & \text{Filmsymbol}_3 & \text{Filmsymbol}_4 & \text{Filmsymbol}_5 & \text{Filmsymbol}_6 \end{matrix} \\ \begin{matrix} \text{Person}_1 \\ \text{Person}_2 \\ \text{Person}_3 \\ \text{Person}_4 \\ \text{Person}_5 \end{matrix} & \begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} \end{pmatrix} \end{matrix}$$

r_{ij}
Zeile i Spalte j

$$U = \begin{matrix} & \begin{matrix} \text{Eigenschaft 1} & \text{Eigenschaft 2} \end{matrix} \\ \begin{matrix} \text{Person}_1 \\ \text{Person}_2 \\ \text{Person}_3 \\ \text{Person}_4 \\ \text{Person}_5 \end{matrix} & \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{pmatrix} \end{matrix}$$
$$M = \begin{matrix} & \begin{matrix} \text{Filmsymbol}_1 & \text{Filmsymbol}_2 & \text{Filmsymbol}_3 & \text{Filmsymbol}_4 & \text{Filmsymbol}_5 & \text{Filmsymbol}_6 \end{matrix} \\ \begin{matrix} \text{Eigenschaft 1} \\ \text{Eigenschaft 2} \end{matrix} & \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} & m_{15} & m_{16} \\ m_{21} & m_{22} & m_{23} & m_{24} & m_{25} & m_{26} \end{pmatrix} \end{matrix}$$

Wichtig: Antwortblatt downloaden und auf deinem PC ausfüllen!

Antwortblatt zu Arbeitsblatt 3

Aufgabe 1 | Zerlegung per Hand berechnen

Wie bist bei der Bestimmung einer Zerlegung vorgegangen? Notiere Beobachtungen aus dem Vergleich mit deinem Gruppenmitglied.

Aufgabe 2 | Bewertung einer gefundenen Zerlegung

Teil a - Vergleich verschiedener Vorhersage-Matrizen

Welche der Zerlegungen ist am besten? Welche ist am schlechtesten? Wie beurteilst du die Güte der Zerlegung?

Fehler der Zerlegung 1:

Fehler der Zerlegung 2:

Fehler der Zerlegung 3:

Teil b - Allgemeine Formel für den Fehler

Stelle eine Formel für den Fehler auf, der sich bei einer Zerlegung der Rating-Matrix R in eine User-Matrix U und eine Movie-Matrix M ergibt.

Formel:

Teil c | Vergleichbarkeit des Fehlers

Kann das Fehlermaß aus Teil b verwendet werden, um vergleichen zu können, welche Vorhersage-Matrix besser ist? Begründe deine Antwort.

Antwortblatt zu Arbeitsblatt 4 (Kurzversion) Langversion auf nächster Seite

Aufgabe 1 | Anwendung des Verfahrens auf größere Ratingmatrizen

Teil a

Beurteile, wie gut die **bekannten** Bewertungen vorhergesagt werden.
Ist die Zerlegung auch für die **unbekannten** Bewertungen gut geeignet?

Teil b | Der Einfluss des Parameters λ

Wie ändern sich die Werte in der Vorhersage-Matrix, wenn du λ größer oder kleiner wählst? Wie ändern sich die Vorhersagen der unbekannten Bewertungen (Nullen in R_1)?

Zusatzaufgabe

Beantworte folgende Fragen:

- Für welchen Wert von λ entspricht die Funktion mit Regularisierungsterm der Funktion ohne Regularisierungsterm?

- Angenommen wir wählen einen großen Wert für λ , zum Beispiel $\lambda = 10$. Wie müssen die Einträge von U und M gewählt werden, damit die Fehlerfunktion mit Regularisierungsterm dennoch einen kleinen Wert liefert?

Wichtig: Antwortblatt downloaden und auf deinem PC ausfüllen!

Antwortblatt zu Arbeitsblatt 4 (Langversion)

Aufgabe 1 | Schrittweise Entwicklung des Optimierungsverfahrens

1. Schritt | Movie-Matrix festlegen

Fehlerfunktion $error_U$:

2. Schritt | Einträge der User-Matrix optimieren

Ableitung bzgl. u_{11} :

Ableitung bzgl. u_{12} :

Platz für das Lösen des linearen Gleichungssystems:

3. und 4. Schritt | User-Matrix festlegen und Einträge der Movie-Matrix optimieren

Fehlerfunktion $error_M$:

Ableitung bzgl. m_{11} :

Ableitung bzgl. m_{12} :

Ableitung bzgl. m_{21} :

Ableitung bzgl. m_{22} :

Platz für das Lösen des linearen Gleichungssystems:

Antwortblatt zu Arbeitsblatt 4 (Langversion)

Aufgabe 2 | Anwendung des Verfahrens auf größere Ratingmatrizen

Teil a |

Beurteile, wie gut die **bekannten** Bewertungen vorhergesagt werden.
Ist die Zerlegung auch für die **unbekannten** Bewertungen gut geeignet?

Teil b | Der Einfluss des Parameters λ

Beantworte folgende Fragen:

- Für welchen Wert von λ entspricht die Funktion mit Regularisierungsterm der Funktion ohne Regularisierungsterm?

- Angenommen wir wählen einen großen Wert für λ , zum Beispiel $\lambda = 10$. Wie müssen die Einträge von U und M gewählt werden, damit die Fehlerfunktion mit Regularisierungsterm dennoch einen kleinen Wert liefert?

Wichtig: Antwortblatt downloaden und auf deinem PC ausfüllen!

Antwortblatt zu Arbeitsblatt 5

Aufgabe 2 | Parameterstudie

features	lambda	Fehler auf den Trainingsdaten	Fehler auf den Testdaten
8	0.00001		
8	0.001		
8	0.01		
8	0.1		
8	0.2		
8	0.5		
8	1		
8	5		
8	10		
8	15		
4	0.3		
6	0.3		
12	0.3		
20	0.3		
40	0.3		
Platz für weitere Ergebnisse			

Einfluss des Parameters λ auf den Trainings- und den Testfehler:

Einfluss des Parameters features auf den Trainings- und den Testfehler:

Zusatzaufgabe 1 | Cold Start Problem

Unter dem Cold Start Problem versteht man, dass Nutzer, die sich neu bei Netflix anmelden noch gar keine Bewertungen abgegeben haben.

- *Wie können wir diesen Nutzern dennoch Filme vorschlagen? Notiert erste Ideen.*

Zusatzaufgabe 2 | Andere Modellierungsansätze

- *Überlegt, wie man Ähnlichkeiten zwischen verschiedenen Usern ausgehend von der Rating-Matrix messen kann!*

Wichtig: Antwortblatt downloaden und auf deinem PC ausfüllen!