

Mining massive Datasets WS 2017/18

Problem Set 4

Rudolf Chrispens, Marvin, Daniela Schacherer

November 20, 2017

Exercise 01

We use the formula $\cos\phi = \frac{a*b}{||a||*||b||}$ where ϕ is the angle between the vectors a and b . The weighting vector w which is multiplied to a and b before calculating the cosine angle is $\begin{pmatrix} 1 \\ \alpha \\ \beta \end{pmatrix}$.

a) Here we have $\alpha = 1$ and $\beta = 1$. We receive the following cosine angles, which indicate that all three vectors point in almost the same direction:

- $\phi_{AB} = 0.13^\circ$
- $\phi_{AC} = 0.17^\circ$
- $\phi_{BC} = 0.28^\circ$

b) Here we have $\alpha = 0.01$ and $\beta = 0.5$. The weighted vectors are thus

$$\begin{array}{l} PS \\ DS \\ MMS \end{array} \begin{pmatrix} A & B & C \\ 3.06 & 2.68 & 2.92 \\ 5 & 3.2 & 6.4 \\ 3 & 2 & 3 \end{pmatrix}$$

We receive the following cosine angles:

- $\phi_{AB} = 7.74^\circ$
- $\phi_{AC} = 7.45^\circ$
- $\phi_{BC} = 14.26^\circ$

c) If we want to select α and β as the invers proportional of the average in the respective component we receive $\alpha = \frac{1}{\frac{500+320+640}{3}} = \frac{1}{487}$ and $\beta = \frac{1}{\frac{6+4+6}{3}} = \frac{1}{5.34}$.

$$\begin{array}{l} PS \\ DS \\ MMS \end{array} \begin{pmatrix} A & B & C \\ 3.06 & 2.68 & 2.92 \\ 1.03 & 0.66 & 1.31 \\ 1.12 & 0.75 & 1.12 \end{pmatrix}$$

With possible rounding errors during the calculation we receive for the angles:

- $\phi_{AB} = 6.01^\circ$
- $\phi_{AC} = 5.25^\circ$
- $\phi_{BC} = 10.67^\circ$

Exercise 01

Consider a web shop that sells furniture and uses a recommendation system. When a new user creates an account and likes one product, he will be presented with similar products on his next visit.

How can a competitor - in principle - try to steal the valuable data for recommendation from this website?

1. -

Does this work better when the web shop implemented a content- based or a collaborative filtering system?

1. -

What data would the competitor be able to infer?

1. -

Would this technique have an impact on the recommendation system, i.e., would this attack create a bias on the data?

1. -

Why is this attack probably not viable in any case?

1. -