

<u>Unit 2 Nonlinear Classification</u>, <u>Linear regression, Collaborative</u>

<u>Course</u> > <u>Filtering (2 weeks)</u>

3. K-Nearest Neighbor Method

> Lecture 7. Recommender Systems >

Audit Access Expires May 11, 2020

You lose all access to this course, including your progress, on May 11, 2020. Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site. **Upgrade now**

3. K-Nearest Neighbor Method K-Nearest Neighbor Method





Video

Download video file

Transcripts

<u>Download SubRip (.srt) file</u> Download Text (.txt) file

Our goal in the movie recommender system problem is to predict the movie ranking that a user would give on a movie that (s)he has not seen.

Let m be the number of movies and n the number of users. The ranking Y_{ai} of a movie $i\in\{1,\ldots,m\}$ by a user $a\in\{1,\ldots,n\}$ may already exist or not. Our goal is to predict Y_{ai} in the case when Y_{ai} does not exist.

$oldsymbol{K}$ -Nearest Neighbour

The K-Nearest Neighbor method makes use of ratings by K other "similar" users when predicting Y_{ai} .

Let $\mathrm{KNN}\left(a\right)$ be the set of K users "similar to" user a, and let $\mathrm{sim}\left(a,b\right)$ be a **similarity measure** between users a and $b\in\mathrm{KNN}\left(a\right)$. The K-Nearest Neighbor method predicts a ranking Y_{ai} to be :

$$\widehat{{Y}}_{ai} = rac{\displaystyle\sum_{b \in ext{KNN}(a)} \sin \left(a,b
ight) Y_{bi}}{\displaystyle\sum_{b \in ext{KNN}(a)} \sin \left(a,b
ight)}.$$

The similarity measure $\sin{(a,b)}$ could be any distance function between the feature vectors x_a and x_b of users a and b, e.g. the euclidean distance $\|x_a - x_b\|$ and the cosine similarity $\cos{\theta} = \frac{x_a \cdot x_b}{\|x_a\| \ \|x_b\|}$. We will use these similarity measures again in *Unit 4 Unsupervised Learning*.

A drawback of this method is that the success of the K-Nearest Neighbor method depends heavily on the choice of the similarity measure. In the next section, we will discuss collaborative filtering, which will free us from the need to define a good

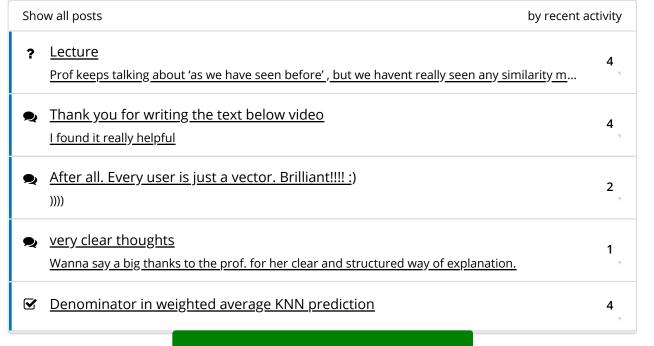
similarity measure.

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 7. Recommender Systems / 3. K-Nearest Neighbor Method

Add a Post



Learn About Verified Certificates

© All Rights Reserved