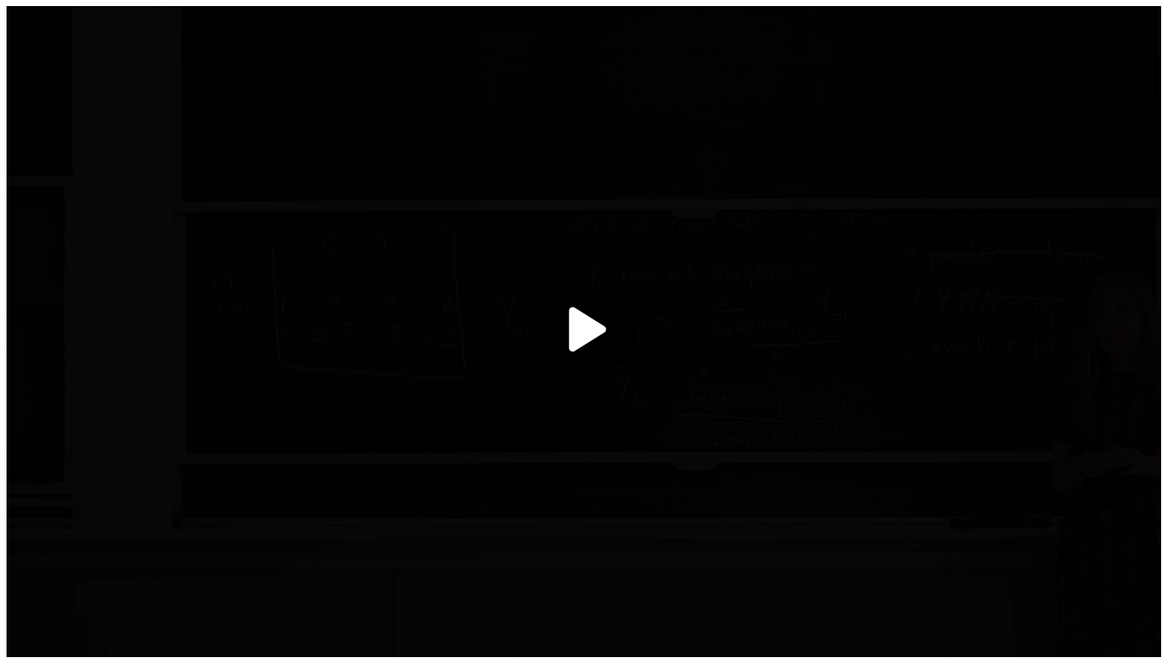
3. K-Nearest Neighbor Method
K-Nearest Neighbor Method



7:30 / 7:30 Speed 1.50x

both in terms of products and in terms of users.

So then you don't have to engineer very sophisticated

similarity measure, your algorithm

would be able to pick up these very complex dependencies.

And for us, as human, it would be definitely not tractable

to come up with such measures.

So now, we are ready to start talking

about the next approach, which is called matrix factorization.

Video
Download video file

Transcripts
Download SubRip (.srt) file

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.

K-Nearest Neighbour

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

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

$$\widehat{Y}_{ai} = \frac{\sum_{b \in \mathrm{KNN}(a)} \mathrm{sim}(a, b)\, Y_{bi}}{\sum_{b \in \mathrm{KNN}(a)} \mathrm{sim}(a, b)}.$$

The similarity measure $\mathrm{sim}(a, b)$ 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

Show all posts ▼ by recent activity ▼

? How do we determine the members of KNN(a)? 2

Similarity measure vs distance 3

? The sentence is harder to read due to exchanging the order of the m, n mention 2

Very clear 2

Learn About Verified Certificates