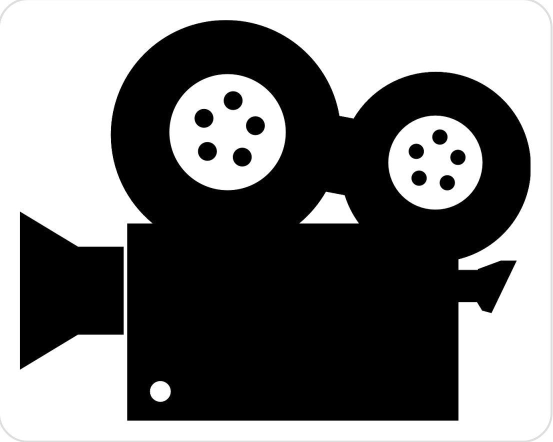
cs110_lab2_als_prediction

databricks
Home
Workspace
Recent
Tables
Clusters
Jobs
Search

Predicting Movie Ratings



One of the most common uses of big data is to predict what users want. This allows Google to show you relevant ads, Amazon to recommend relevant products, and Netflix to recommend movies that you might like. This lab will demonstrate how we can use Apache Spark to recommend movies to a user. We will start with some basic techniques, and then use the Spark ML (https://spark.apache.org/docs/1.6.2/api/python/pyspark.ml.html) library's Alternating Least Squares method to make more sophisticated predictions.

For this lab, we will use a subset dataset of 20 million ratings. This dataset is pre-mounted on Databricks and is from the MovieLens stable benchmark rating dataset (http://grouplens.org/datasets/movielens/). However, the same code you write will also work on the full dataset (though running with the full dataset on Community Edition is likely to take quite a long time).

Send feedback