Movie recommendations using Spark’s machine learning library

Distributed Machine Learning with MLLib

Recommendations using Alternating Least Squares

Vector (dense or sparse) – To concise the dat. Let’s say we have a dataset of all the movies made and the ratings given by a person, the actual number of ratings given by a person is less, we can have this data in our vector.

LabeledPoint – Supervised machine learning – predict a label for a point

Rating is used by MLLib

Concept of ALS:

Decomposition of large user/item matrix into lower dimensional user factors and item factors and estimate the rating by multiplying those factors according to the equations. Similar to factor analysis.