Bayesian Personalized Ranking (BPR) with Spark

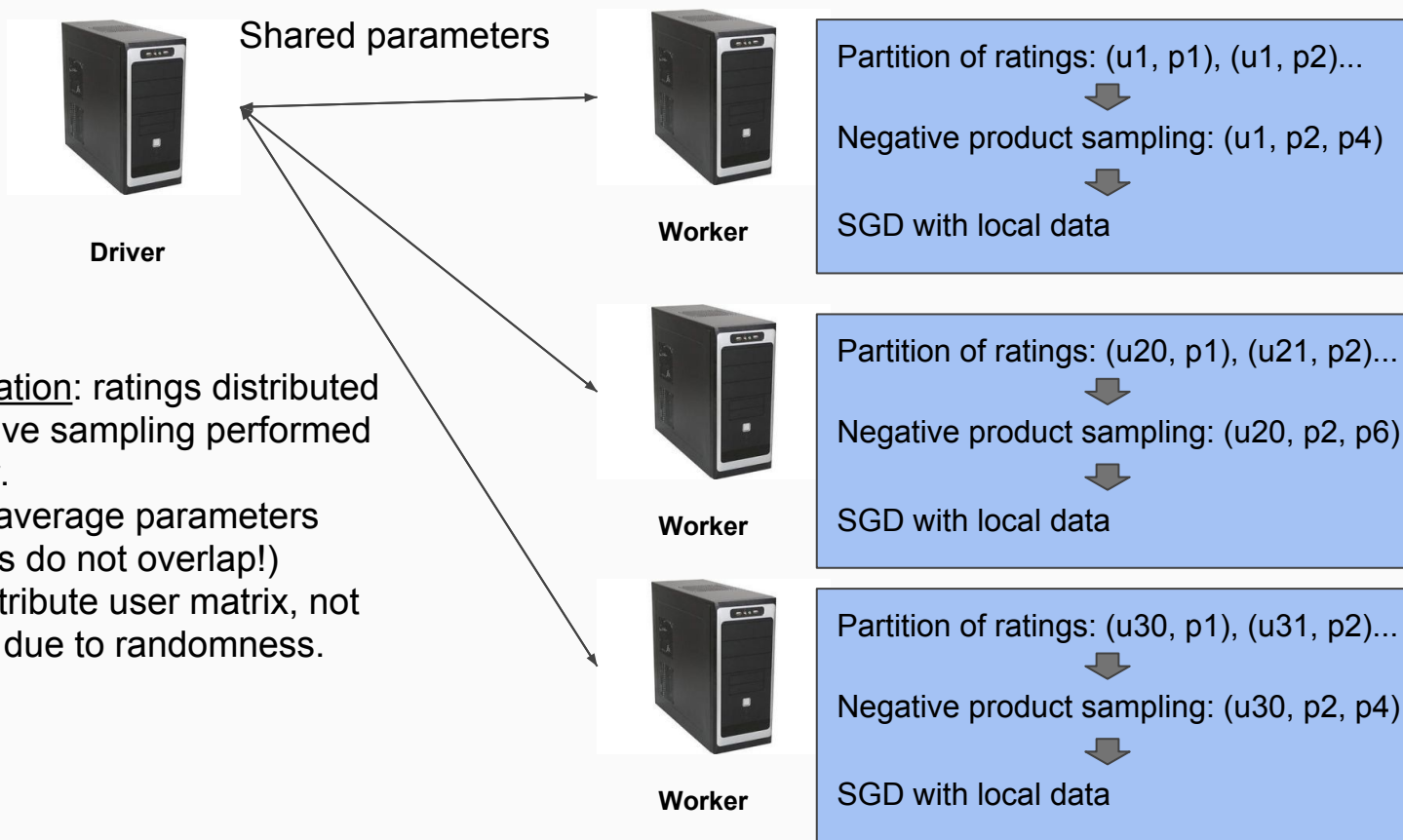Alfredo Láinez Rodrigo *and* Luke de Oliveira

Bayesian Personalized Ranking for Implicit Feedback (BPR)

- Bayesian framework for personalized recommendations.
- Abstract away model from recommender system estimation

$$p(\Theta|>_u) \sim p(>_u|\Theta)p(\Theta)$$

- Implicit feedback $\rightarrow$ maximize AUC w.r.t. negative samples
- BPR expressed as optimization problem $\rightarrow$ SGD
- Learn ordering $>_u$ over (i, j) items
- Chosen underlying model: Matrix Factorization

Parallelization in Spark

Shared parameters

Driver

Worker

Partition of ratings: (u1, p1), (u1, p2)...

Negative product sampling: (u1, p2, p4)

SGD with local data

Worker

Partition of ratings: (u20, p1), (u21, p2)...

Negative product sampling: (u20, p2, p6)

SGD with local data

Worker

Partition of ratings: (u30, p1), (u31, p2)...

Negative product sampling: (u30, p2, p4)

SGD with local data

- <u>Data parallelization</u>: ratings distributed by user, negative sampling performed in each worker.
- <u>Parallel SGD</u>: average parameters (note that users do not overlap!)
- Possible to distribute user matrix, not product matrix due to randomness.