Streaming Inference for Infinite Feature Models

Rylan Schaeffer[1,2], Yilun Du[3], Gabrielle Kaili-May Liu[2], Ila Rani Fiete[2,4]
[1]Stanford CS, [2]MIT BCS, [3]MIT EECS, [4]McGovern Institute for Brain Research

Introduction

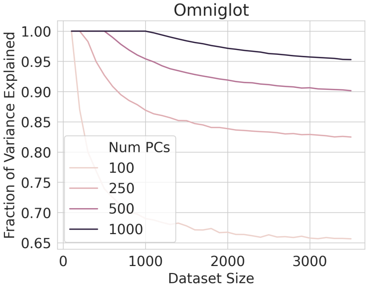
- Biological intelligence operates in a radically different data regime than (most) artificial intelligence
- Biological intelligence must contend with data that is (i) unsupervised, (ii) streaming (i.e. online) and (iii) non-stationary
- In this data regime, how should one approach learning?
- In this work, we set aside the non-stationary aspect
- We specifically ask how to use features models (a widely-used family of unsupervised learning algorithms that includes PCA, FA, ICA, NMF) on streaming data

Motivation

On streaming data, using a preset fixed number of features results in an inability to model the data!

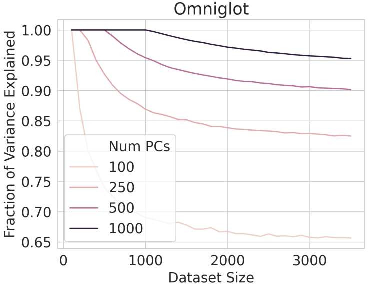
Motivation

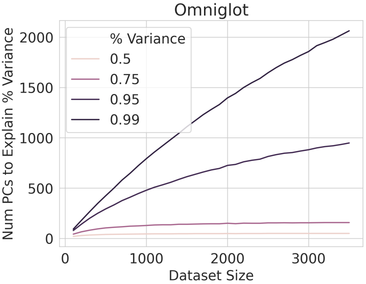
On streaming data, using a preset fixed number of features results in an inability to model the data!



Motivation

On streaming data, using a preset fixed number of features results in an inability to model the data!





How can one define a learning algorithm that grows in representational capacity as necessitated by the stream of data?

Approach: The Recursive Indian Buffet Process

- Enable any feature model to add features, as necessitated by the data, via Bayesian nonparametrics

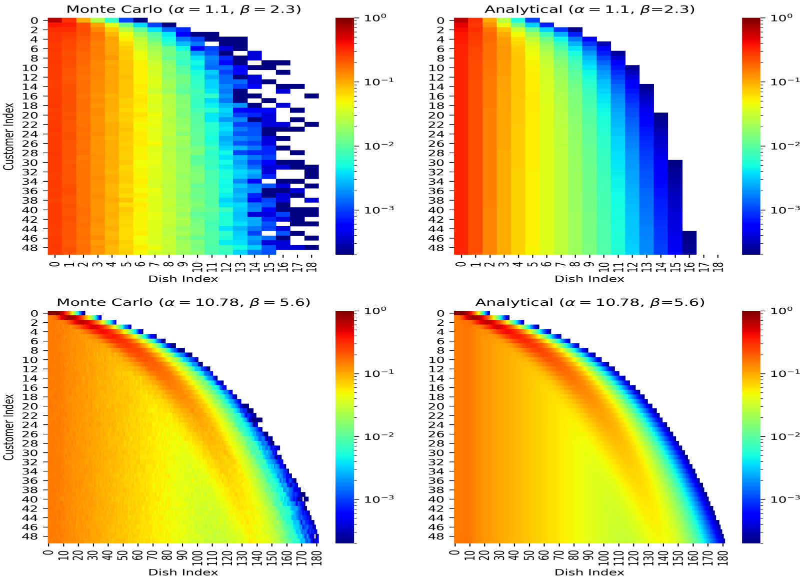
Approach: The Recursive Indian Buffet Process

- Enable any feature model to add features, as necessitated by the data, via Bayesian nonparametrics
- Specifically, we derive a novel recursive form of the Indian Buffet Process designed specifically for streaming data:

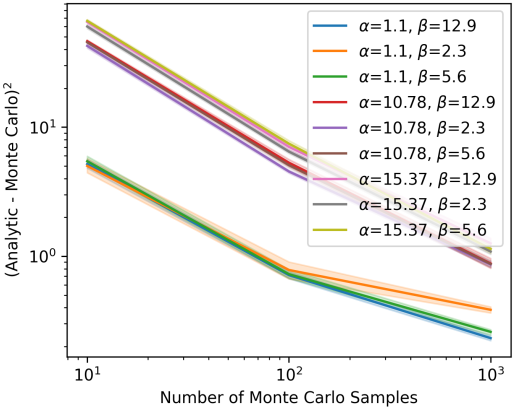
$$\underbrace{p(z_{nk}=1)}_{\text{P(n obs contains kth feature)}} = \underbrace{\frac{1}{\beta+n-1}\sum_{n'<n} p(z_{n'k}=1)}_{\text{How probable is kth feature}}$$

$$+ \underbrace{p(\Lambda_{n-1} \leq k-1) - p(\Lambda_{n-1}+\lambda_n \leq k-1)}_{\text{Pressure to create new features as $n \to \infty$}}$$

Monte Carlo Estimate vs. Analytical Expression



Monte Carlo Estimate vs. Analytical Expression



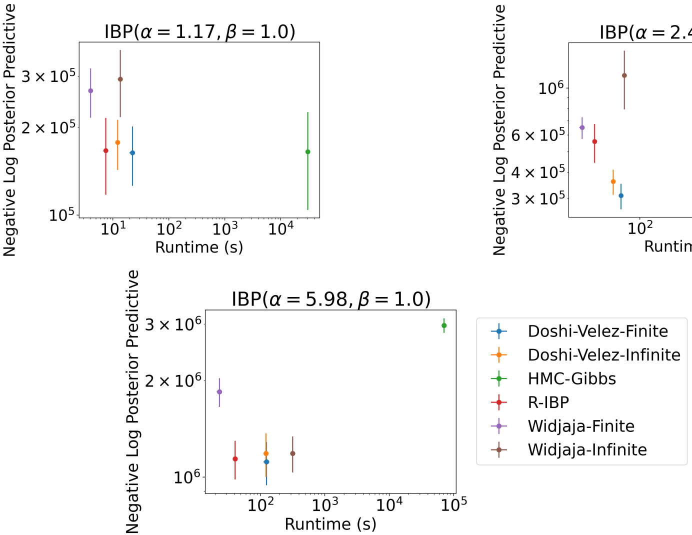
As the number of Monte Carlo samples increase, the average difference between the analytical R-IBP and the numerical estimate falls as an approximate power law for all combinations of α, β.

Intuition: Analytical Results

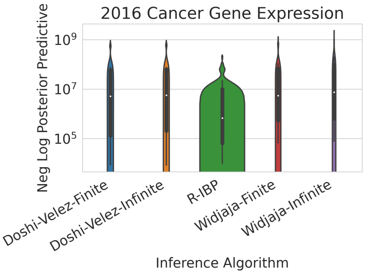
- The Recursive IBP follows a simple objective function: learn features A and indicators Z that maximize the likelihood while penalizing the number of used features Λ_N

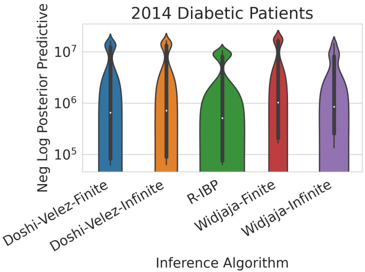
$$\underset{A,Z,\Lambda_N}{\arg\min} \quad \text{Tr}\left[(O - ZA)^T(O - ZA)\right] + \gamma^2\Lambda_N$$

Synthetic Data: Faster & Equal/Better Performance



Tabular Data: Faster & Equal/Better Performance





On cancer gene expression (left) and diabetic patient (right) data, R-IBP matches or outperforms baseline algorithms across hyperparameter configurations.

Conjecture

The Recursive IBP outperforms non-streaming offline baselines because its graphical structure avoids chain multiplying inferred variables, instead adding inferred variables to running sums.