Restricting exchangeable nonparametric distributions

Sinead A. Williamson Steven N. MacEachern Eric P. Xing

THE UNIVERSITY OF TEXAS AT AUSTIN

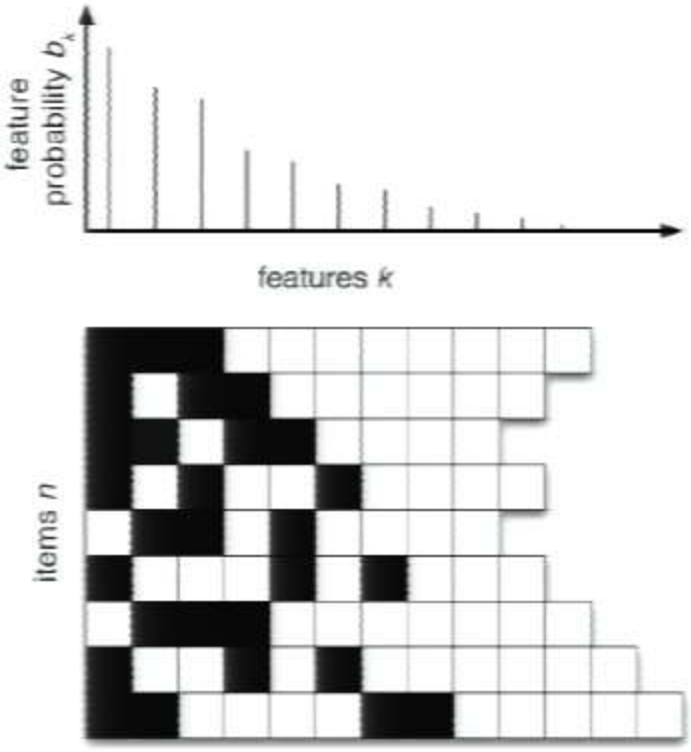
THE OHIO STATE UNIVERSITY

Carnegie Mellon University

Indian buffet process

THE UNIVERSITY OF TEXAS AT AUSTIN

- Distribution over binary matrices with infinitely many columns and exchangeable rows.
- Idea: Rows = data points; columns = features.
- Feature probabilities distributed according to a *beta process*.
- Each row distributed according to a *Bernoulli process*.
- Number of ones per row marginally **Poisson**-distributed.



Restricting the number of features per data point

THE UNIVERSITY OF TEXAS AT AUSTIN

- Poisson number of features per data point may not be appropriate:
  - We may know the number of features per data point – e.g. speakers in a dialogue, members of a team.
  - We may believe the number of features exhibits power-law behaviour.
- Poisson behaviour results from the beta process and the Bernoulli process being *completely random measures*.
- We can remove this behaviour by directly restricting the support of the Bernoulli process, e.g.
  - Bernoulli process restricted to have K non-zero entries $\rightarrow$ binary matrix with K non-zero entries/row.
  - Bernoulli process restricted to have a chosen distribution f over the number of non-zero entries $\rightarrow$ binary matrix with $n_i \sim f$ entries/row.
- Resulting matrix is still exchangeable.

More flexible distributions, more flexible models

THE UNIVERSITY OF TEXAS AT AUSTIN

- More flexible distributions over matrices:

- When the number of features is known, IBP restricted to have K features recovers better representations than restricted IBP.
- In a classification task on text data, IBP restricted to have heavier-tailed marginals performs better than unrestricted IBP.
- For experimental details and more information, visit poster **Fri38**.