

Iterative Latent Spaces for Interpretable High-Dimensional Modeling

Jackson Loper, David Blei, and Liam Paninski
Columbia University

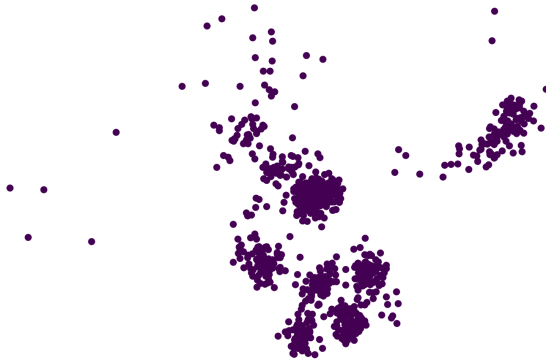
Iterative clustering is a natural idea



Iterative clustering is a natural idea



Iterative clustering is a natural idea



Iterative clustering is a natural idea

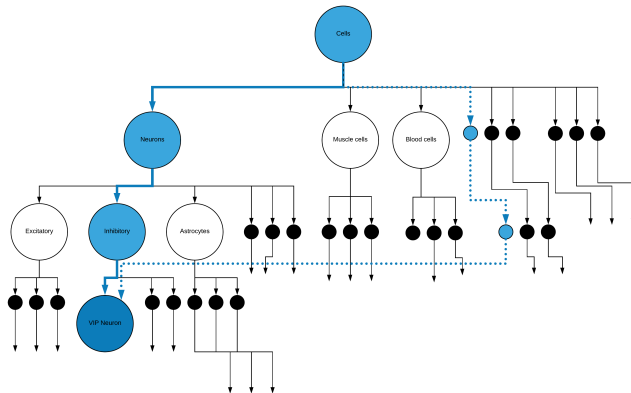
No two-dimensional space can smoothly represent the full diversity of cells.

Latent-variable generative modelling is a natural idea

Especially natural for multi-modality data.

$$\mathbb{P}(X_1, X_2, X_3, \dots) = \int \left(\left(\prod_i \mathbb{P}(X_i|Z) \right) \mu(dZ) \right)$$

Put them together



Empirical results

Tasic, Bosiljka, Zizhen Yao, Kimberly A. Smith, Lucas Graybuck, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy et al. “Shared and distinct transcriptomic cell types across neocortical areas.” bioRxiv (2017): 229542.

Empirical results

- Amortized variational inference
- Zero-inflated negative binomial distribution for gene expression

Empirical results

- Amortized variational inference
- Zero-inflated negative binomial distribution for gene expression

Method	Train	Test
Factor analysis $k = 16$	-4.2	-4.2
Factor analysis $k = 64$	-4.1	-4.3
HNBPF $k = 2$	-5.7	-5.7
HNBPF $k = 16$	-3.8	-3.9
HNBPF $k = 64$	-3.8	-5.7
ZINB-NN $k = 2$	-3.6	-3.6
ILS $k = 2 + 2$	-3.5	-3.5

More flexible, but still interpretable

