

Differential Expression Analysis Techniques for Single-Cell RNA-seq Experiments

for the Computational Biology Doctoral Seminar (CMPBIO 293),
organized by N. Yosef & T. Ashuach, Spring 2018, UC Berkeley

Kevin Benac and Nima Hejazi

Group in Biostatistics,
University of California, Berkeley

11 April 2018

Outline

Introduction

- Data (Kevin)

- Objective (Nima)

Methodology

- ZINB-WaVE (Kevin)

- DropLasso (Nima)

- Comparison

Conclusions

- Review

The Data: Single-Cell RNA-seq



The Data: Single-Cell RNA-seq



The Data: Single-Cell RNA-seq



Outline

Introduction

Data (Kevin)

Objective (Nima)

Methodology

ZINB-WaVE (Kevin)

DropLasso (Nima)

Comparison

Conclusions

Review

The Objective: Differential Expression

- ▶ Many experimental settings seek to isolate a subset of biomarkers from the full (larger) observed list observed in the experiments.
- ▶ Why “differential”? Finding a subset of relevant biomarkers is done with respect to a particular condition of interest (e.g., disease).
- ▶ ...

The Objective: Differential Expression

- Regularized Linear Models:

$$\min_{w \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^n \mathcal{L}(w, x_i, y_i) + \lambda \Omega(w) \right\}$$

- Lasso for continuous outcomes (squared-error loss):

$$\min_{w \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^d w_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^d |w_j| \right\}$$

Outline

Introduction

Data (Kevin)

Objective (Nima)

Methodology

ZINB-WaVE (Kevin)

DropLasso (Nima)

Comparison

Conclusions

Review

ZINB-WaVE I

► ...

► ...

ZINB-WaVE II

► ...

► ...

ZINB-WaVE III

► ...

► ...

ZINB-WaVE IV

► ...

► ...

ZINB-WaVE V

► ...

► ...

Outline

Introduction

Data (Kevin)

Objective (Nima)

Methodology

ZINB-WaVE (Kevin)

DropLasso (Nima)

Comparison

Conclusions

Review

DropLasso I

- ▶ Introducing dropout ($\delta_i \sim \text{Bern}(p)^d$):

$$\min_{w \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\delta_i} \mathcal{L} \left(w, \delta_i \odot \frac{x_i}{p}, y_i \right) + \lambda \|w\|_1 \right\}$$

- ▶ ...

$$\begin{aligned} \mathbb{E}_{\delta_i} \sum_{j=1}^d w_j \left(\delta_i \odot \frac{x_i}{p} \right)_j &= \sum_{j=1}^d \mathbb{E}_{\delta_i} w_j \delta_{i,j} \frac{x_{i,j}}{p} \\ &= \sum_{j=1}^d w_j x_{i,j} \end{aligned}$$

DropLasso II

► ...

► ...

DropLasso III

► ...

► ...

DropLasso IV

► ...

► ...

DropLasso V

► ...

► ...

Outline

Introduction

Data (Kevin)

Objective (Nima)

Methodology

ZINB-WaVE (Kevin)

DropLasso (Nima)

Comparison

Conclusions

Review

ZINB-WaVE v. DropLasso I

► ...

► ...

ZINB-WaVE v. DropLasso II

► ...

► ...

ZINB-WaVE v. DropLasso III

► ...

► ...

Outline

Introduction

- Data (Kevin)

- Objective (Nima)

Methodology

- ZINB-WaVE (Kevin)

- DropLasso (Nima)

- Comparison

Conclusions

- Review

Review



References I

- Beyrem Khalfaoui and Jean-Philippe Vert. DropLasso: A robust variant of Lasso for single-cell RNA-seq data. *arXiv preprint arXiv:1802.09381*, 2018.
- Davide Risso, Fanny Perraudeau, Svetlana Gribkova, Sandrine Dudoit, and Jean-Philippe Vert. ZINB-WaVE: A general and flexible method for signal extraction from single-cell RNA-seq data. *bioRxiv*, 2017.