# "Estimating the number of clusters using cross-validation" Summary of changes to the revised manuscript

Wei Fu          Patrick O. Perry

January 10, 2019

We thank the editors and reviewers for their time reading our submission, and for providing many helpful suggestions for improvement. We also thank you for the opportunity to revise and resubmit our manuscript. We have revised our manuscript carefully taking into account your comments.

Detailed responses to reviewer comments follow below.

## Response to Editor

- *Fix the figure captions.*

  **Response:** Per your instruction, all figure captions now include (1) what is the plot about, (2) specific details of plot, like what type of display and how variables are mapped, (3) the most important thing that the reader should learn.

  TODO

- *Cite R packages.*

  **Response:** We added citations for all packages used in our work: TODO

- *Reprodicibility.*

  **Response:** TODO

## Response to Associate Editor

- *Compare to the slope method*

  TODO

## Response to Reviewer 1

(No further comments.)

# Response to Reviewer 2

- *Bottom of page 4. The authors need to clarify the notation*

  **Response:** <mark>TODO</mark>

- *Page 6. Clarify that $n \gg K$, $m \gg L$.*

  **Response:** <mark>TODO</mark>

- *Section 2.2. Clarify.*

  **Response:** <mark>TODO</mark>

- *Section 3. Note that slope is self consistent.*

  **Response:** <mark>TODO</mark>

- *Proposition 1 improvements.*

  **Response:** <mark>TODO</mark>

  We chose to keep the proof in the main text, but if there are space constraints we will take your suggestion and move the proof to an appendix.

- Proposition 2 improvements.

  **Response:** <mark>TODO</mark>

- *Page 8, line 14. Generalization.*

  **Response:** <mark>TODO</mark>

  We clarified what we meant by saying the proof "generalizes to non-normal distributions easily". You rightly point out that it will not generalize to all possible situations, but only to distributions with finite moments and cases where both clusters have the same distribution.

- *Figure 2. Please define correctly the x,y axis*

  **Response:** <mark>TODO</mark>

- *Page 20, first paragraph.*

  **Response:** Your comment was that "Proposition 2 only shows that if $|\rho| < 0.5$ the method works correctly." Did you mean page 14 instead? We think you may have meant that our statement in the first paragraph of page 14, "Proposition 2 tells us that Gabriel cross-validation fails when there is strong correlation between the variables" is too strong.

- *Compare to standardized Euclidean norm k-means.*

  **Response:** <mark>TODO</mark>

- *Compare to slope*

  **Response:** <mark>TODO</mark>