Internship report

Alternative PCA algorithms analysis with missing values

Léopold Guyot

May 21, 2025

Abstract

Principal Component Analysis (PCA) is commonly used in single-cell proteomics (SCP), but most used PCA algorithm don't support missing values, therefore the data require imputation methods that can introduce biases. This report explores PCA methods that work on missing values, specifically NIPALS and PPCA, in SCP. Real SCP data is used to evaluate the impact of missing values on cell arrangement in the PCA space and assess downstream analysis. Results show that NIPALS and PPCA yield comparable PCA results to SVD. Changes in local cell arrangement are observed with increasing missing value rates, but overall patterns remain similar. NIPALS should become the preferred method to perform PCA on SCP data set, this way avoiding imputation drawbacks.

1 Introduction

Start with generic context – Then zoom in: why is your approach necessary and relevant within this context? – Finish with a brief description of your method and what you found out – Read background information! – Brief, 1 page (font size 11, normal margins)

2 Methods

– Describe the datasets you are using • Both their content and where you got them – Describe the methods you implemented to get your results • Describe separate steps required (reproducibility!) • Overview figures can help. – 2-3 pages

3 Results

– Describe your results • This can be in relation to: – Different input/validation data – Different method parameterisations, ... – Explain what the results mean – Use tables for numbers (do not list in the text) – Figures for distributions, relationships, ... (easier to understand than text) – 2-3 pages

4 Discussion

Which issues did you identify, and which problems did you encounter? – What is different about your approach (and the results you get) in comparison to the original method? Why? – What are advantages/disadvantage of each method? – 1 page

5 Conclusion

References

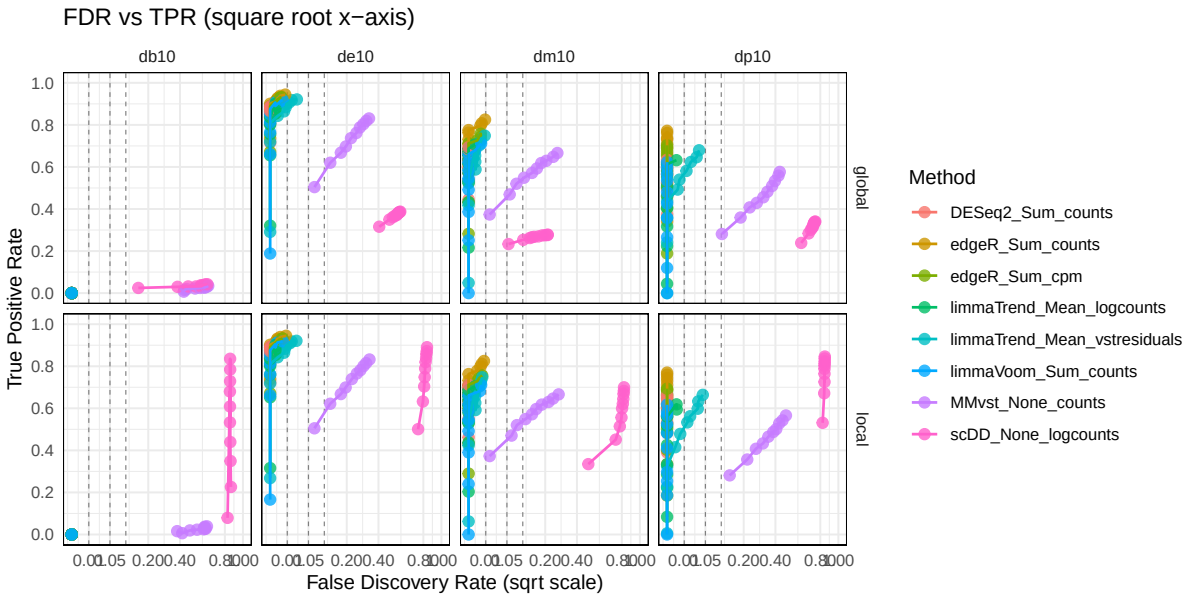

Figure 1:

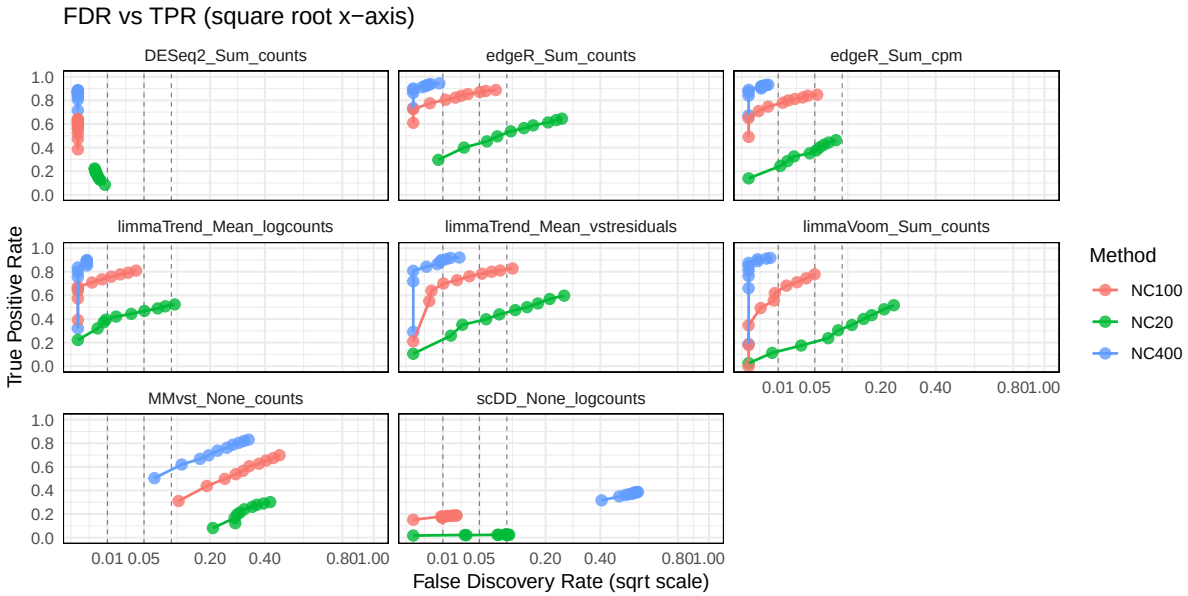

Figure 2: