

Dimension reduction workshop

Alex Diaz-Papkovich
June 25, 2020

Workshop goals

- By the end of the workshop you will:
- Be familiar with the motivation for DR
- Learn some common methods (PCA, t-SNE, UMAP)
- Be able to run code for these methods
- Understand real world applications
- Have some code for future work

Workshop structure

- Jupyter notebooks and R Markdown files
- Will try to to everything online
- Code is available on github
 - https://github.com/diazale/dimension_reduction_workshop
- Should run on your machine pretty easily

Workshop structure

- Basically four parts
 1. Introduction
 2. Playing with methods
 3. Walkthrough with genotype data
 4. Walkthrough with scRNA-seq data

1. Introduction

- What dimension reduction is
- Why we use dimension reduction
- Some underlying ideas and mechanics
- Intuition behind dimension reduction
- Outline of most common methods
- Some cautionary notes
- Examples of application of methods in the field

2. Methods

- Python notebook
- Introduction to PCA, t-SNE, UMAP
- Will generate data and use the iris flower data
- Goals:
 - Get familiar with code to use methods
 - Carry out basic visualizations
 - See some differences between methods

3. Genotype data

- Python notebook using Thousand Genomes data
- Goals:
 - Get familiar with Python code
 - Visualize population structure
 - See differences between different methods
 - Generate interactive plots

4. scRNA-seq data

- Walkthrough of the `Seurat` package in R
- Following the Satija lab's tutorial
- Goals:
 - Get familiar with DR on scRNA-seq data
 - Visualize data (e.g. gene expression)
 - See how DR can fit into scRNA-seq analysis pipeline

Contact

- alex.diaz-papkovich@mail.mcgill.ca
- <https://github.com/diazale>