

STAT 205 Home Work 1

Honglei Ren

January 17, 2020

0.1 Developmental Trajectory Clustering

The developmental trajectories were hierarchically clustered based on their geometric distance in PCA space. More specifically, the *fcluster* method in scikit-learn package was used in hierarchical clustering ([citation?](#)), and the geometric distance between trajectories A and B were defined as the sum of the pair-wised Euclidean distance between two corresponding stages, i.e.

$$\|A - B\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (A_{i,j} - B_{i,j})^2} \quad (1)$$

, where $\|\cdot\|_F$ is the Frobenius norm, A and B are two developmental trajectories represented by m by n matrices, m is the number of developmental stages in single cell data, n is the number of PCA components used in clustering. Each trajectory is a dimension reduced coexpression landscape of a gene pair, and all gene pairs in clustering were downloaded from the gene pair list proposed by ([Briggs et al ref??](#))