Project 2 Report by

- Chua Chin Siang
- Tey Shiwei

Summary

This project was done by variational inference algorithm of LDA. The overall ancestor population assignments are done with variational inference based on each individuals, where only non-zero genotypes were considered in the algorithm. The results for each individuals were concatenated and saved as theta/gamma in our code.

For individual 1, run LDA inference to find for each genotype locus, store it as a matrix of size n1xK, being the indicator function, is the number of non-zero genotypes represent in individual 1. Save it as phi1.out

File saved as phi1.out.npy

Construct a matrix of size M K to represent the ancestor assignments for all individuals in the population. For each individual i, run LDA inference to find gamma, and store it as row of phi (hint: update probabilities in log-space to avoid overflow and underflow issues). Save it as Theta.out

File saved as theta.out.npy

Report the number of iterations and time taken to get to convergence for running inference on all M individuals. (convergence criteria: absolute change in each value of gamma/phi is less than e = 1e-3)

Repeat the experiments for alpha = 0:01; alpha = 1; alpha = 10, and discuss the change in the ancestor population assignments to the individuals , and iterations required for convergence change as alpha change.

	Alpha = 0.01	Alpha = 0.1	Alpha = 1	Alpha = 10
Iterations	6044	5935	3434	1516
Time taken	97.3	92.3	58.5	26.8
Conclusion	● As alpha increases, the number of iterations and time taken reduces ● Convergence criteria : Absolute changes for gamma and phi < 1e-3			
Conclusion for ancestor assignments	As shown in the chart below, ● The ancestor population assignments were far apart (0 to 80) when alpha is small. It follows the prior. ● When alpha value increased to 10, the ancestor population assignments values are approximately within 20 to 60. It follows the mixture of everything.			

Chart of Individual Ancestor Assignments for Different Alpha values