Supplementary materials for "A Hybrid Mixture Approach for Clustering and Characterizing Cancer Data"
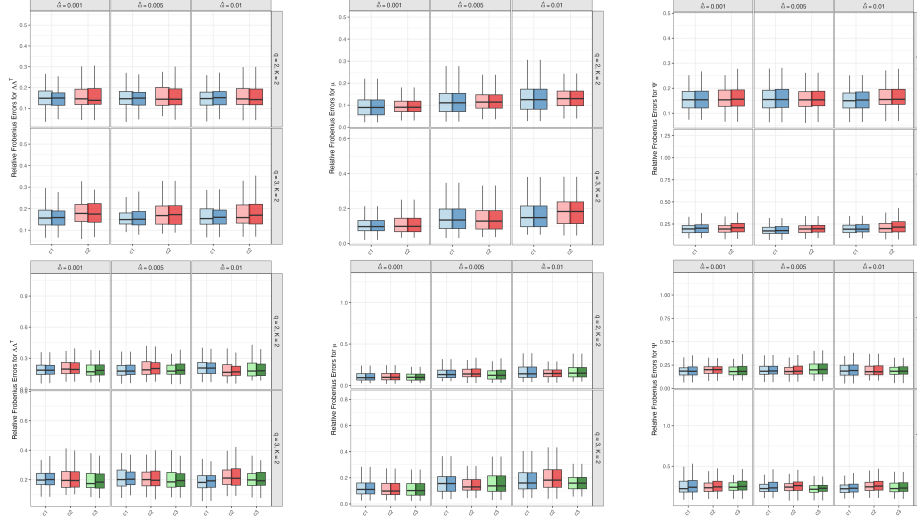
kazeem kareem and Fan Dai

April 2025

Figure 1: Relative Frobenius Errors of Estimated parameters compared to true parameters. Dark shades denote GMMFAD, light shades denote EMMIX.
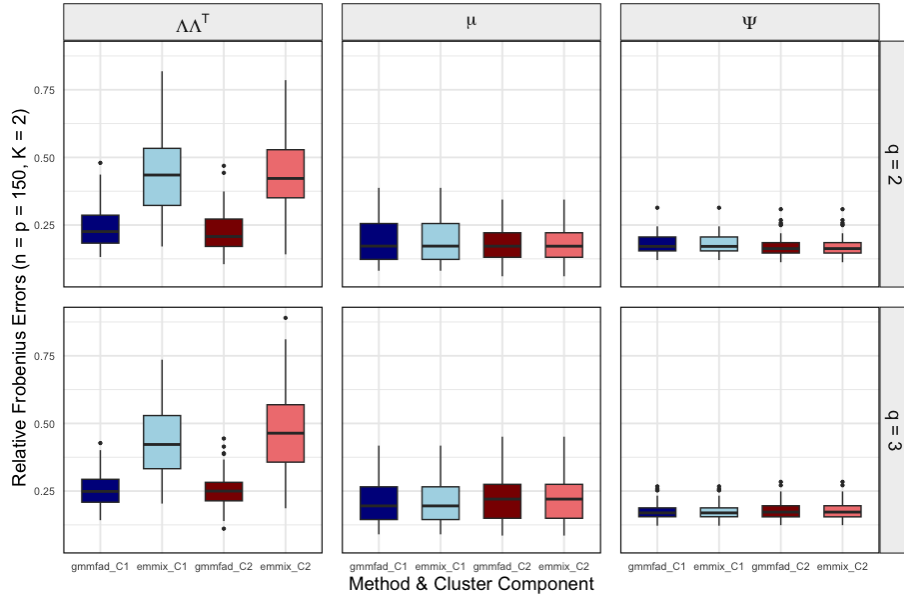

Figure 2: Relative Frobenius errors of parameters for $n = p = 150$, $K = 2$. Dark shades denote GMMFAD, light shades denote EMMIX. Different colors denote different components.
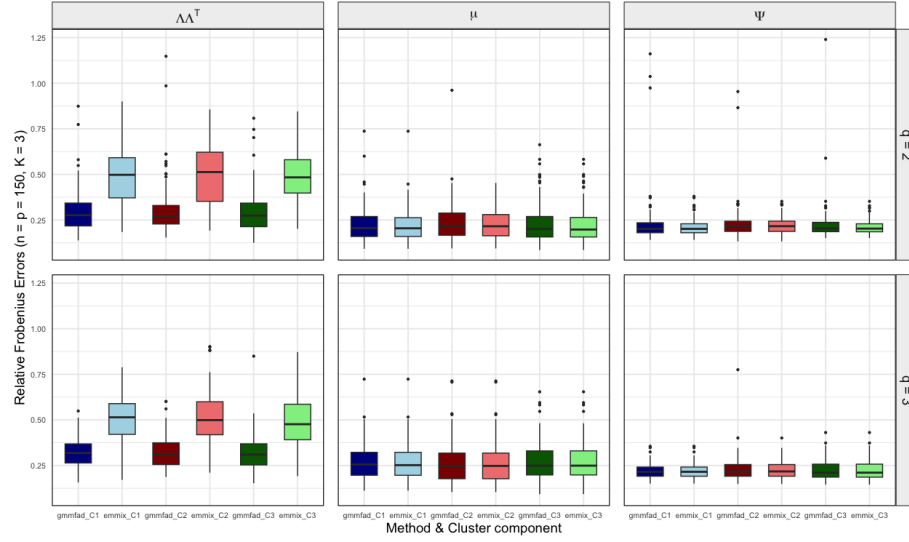
Figure 3: Relative Frobenius errors of parameters for $n = p = 150$, $K = 3$. Dark shades denote GMMFAD, light shades denote EMMIX. Different colors denote different components.
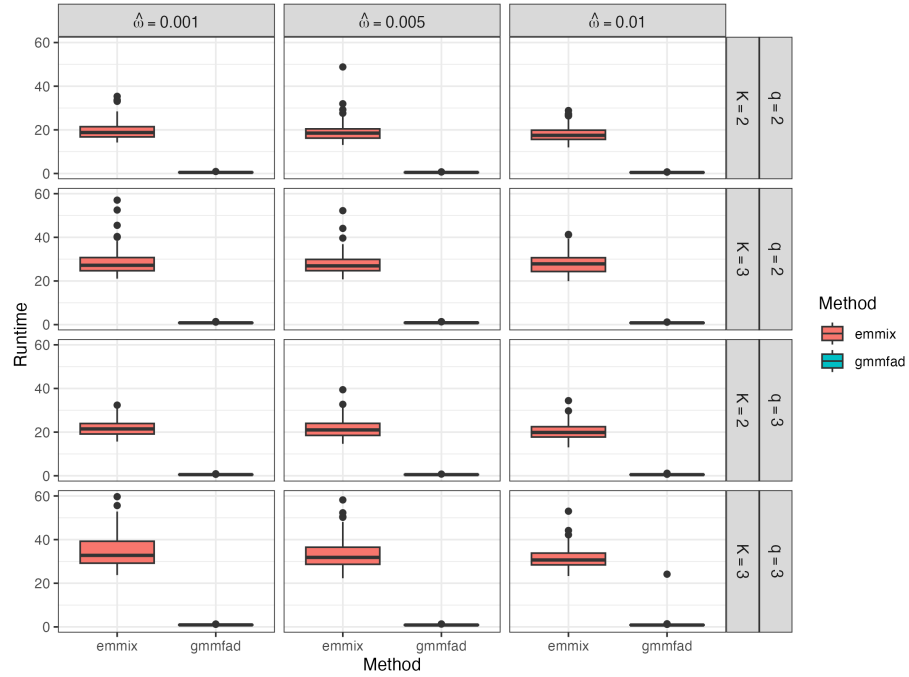

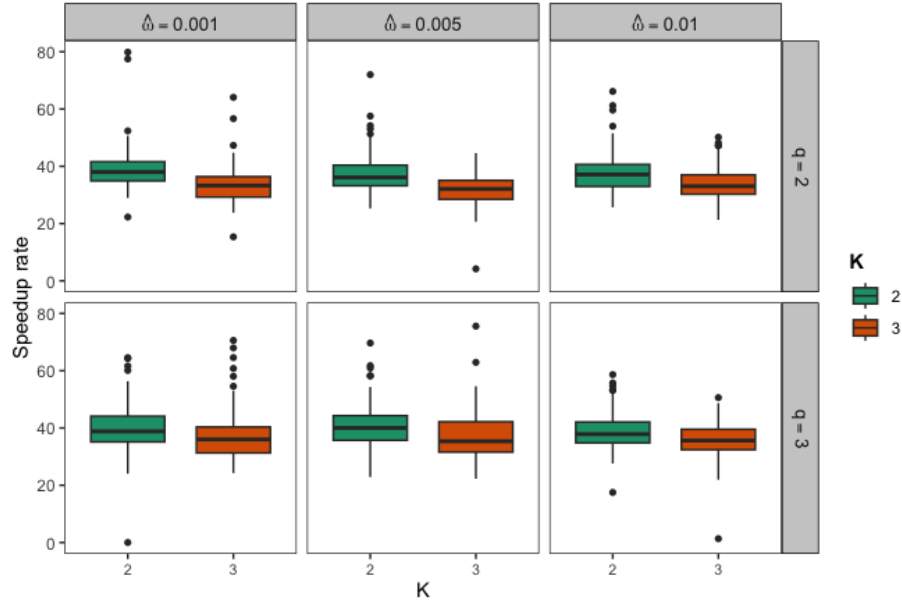Figure 4: CPU times compared for $n = 300$, $p = 10$

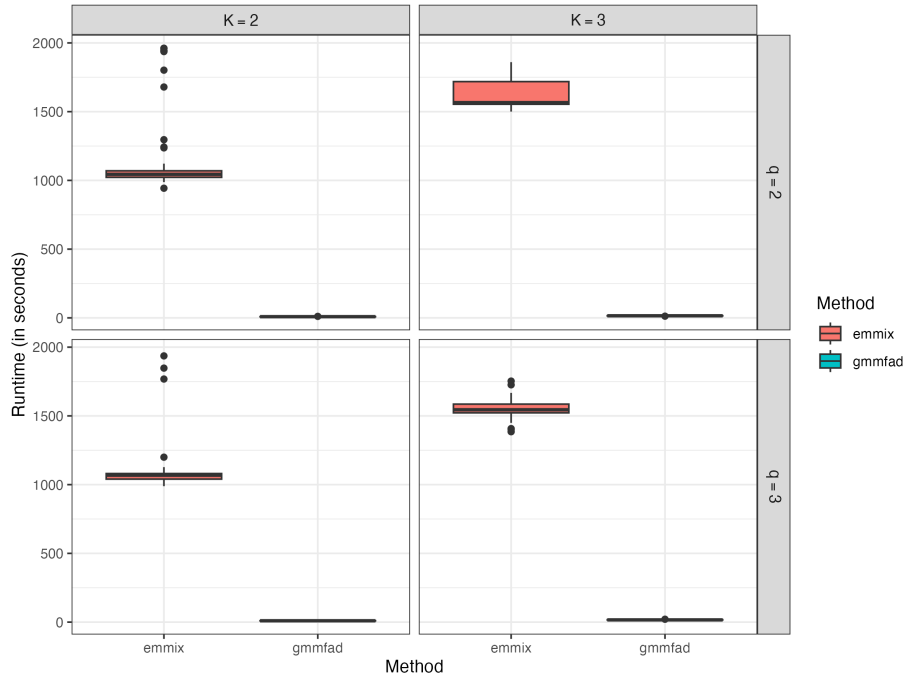Figure 5: Time Speedup: GMMFAD vs EMMIX for $n = 300$, $p = 10$

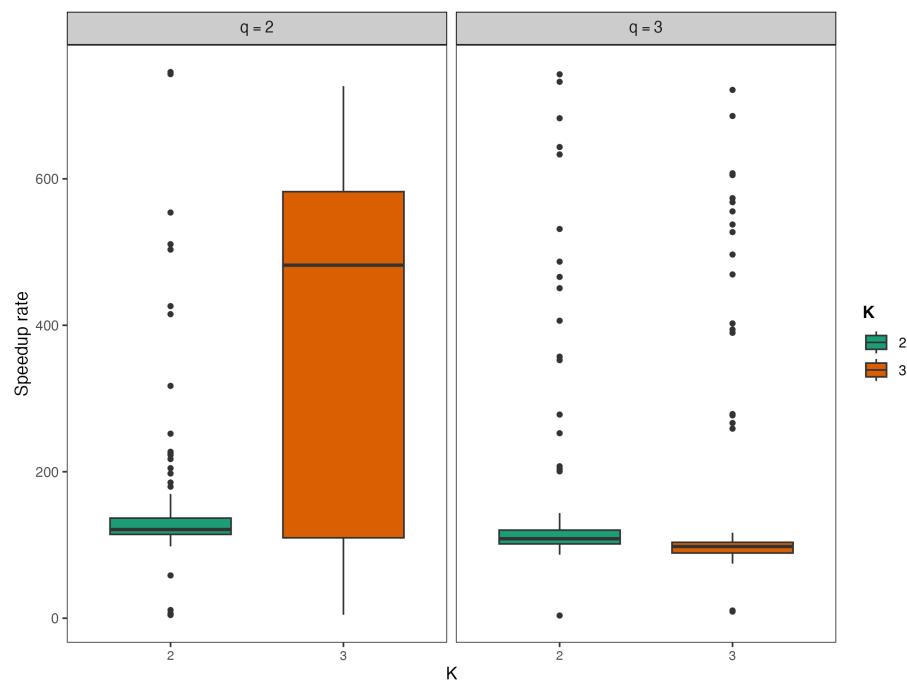Figure 6: Algorithms' runtimes compared for $n = 150$, $p = 150$

4

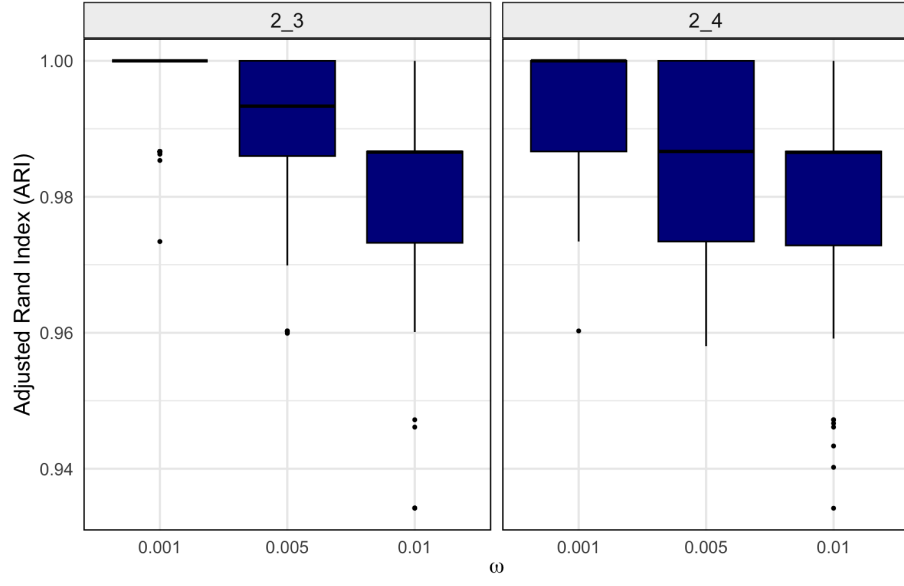Figure 7: Time Speedup of GMMFAD compared to EMMIX for $n = 150$, $p = 150$

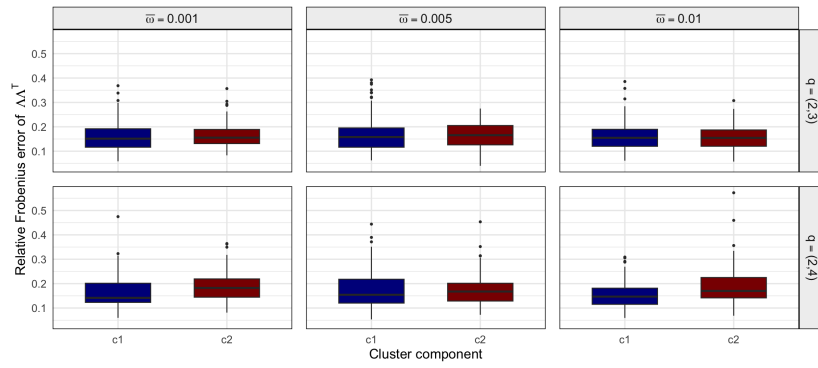Figure 8: Adjusted Rand Indices (ARI) for GMMFADq ($n = 300, p = 10$)

Figure 9: Frobenius Errors of parameters from GMMFADq ($n = 300, p = 10$)
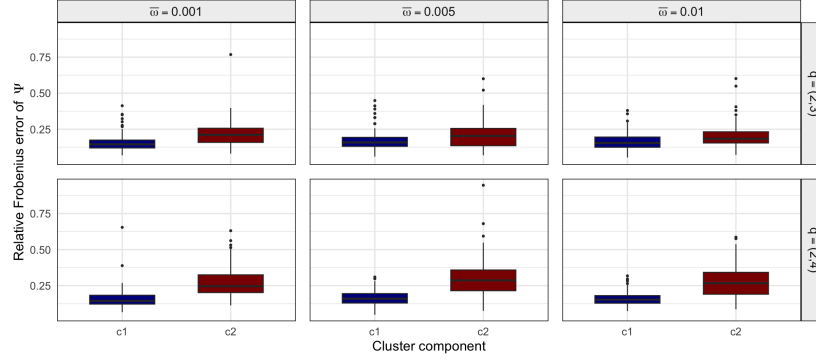
Figure 10: Frobenius Errors of parameters from GMMFADq ($n = 300, p = 10$)
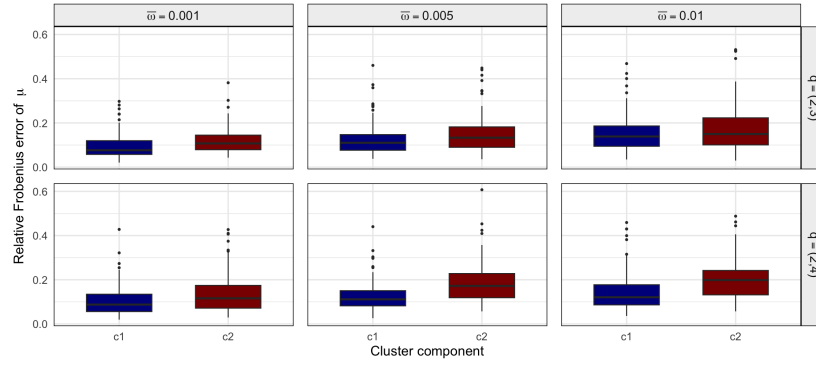

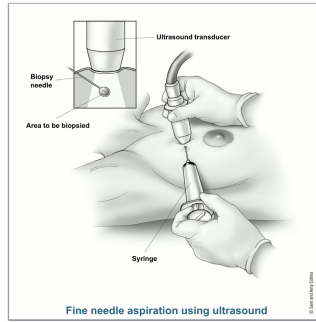Figure 11: Frobenius Errors of parameters from GMMFADq ($n = 300, p = 10$)

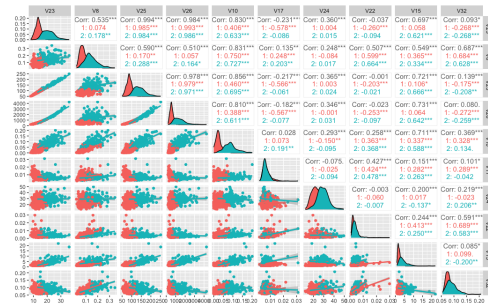Figure 12: Source: American Cancer Society)

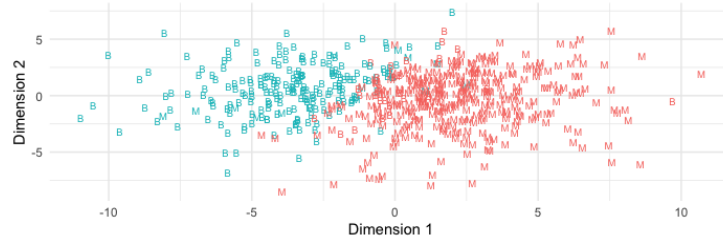Figure 13: Correlation matrix of randomly selected features

Figure 14: Multidimensional Scaling representation of the Wisconsin Breast Cancer data
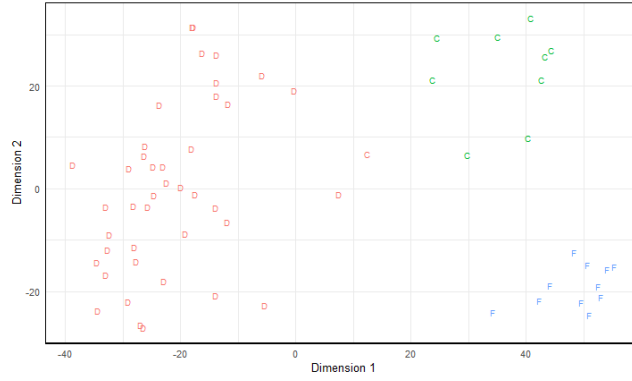

Figure 15: Multidimensional scaling plot of the lymphoma data on two dimensions. D, C and F correspond respectively to DLBCL, CLL and FL.
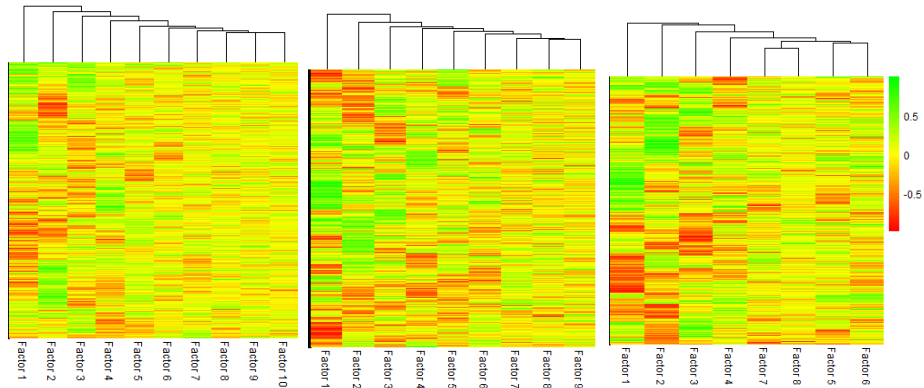

Figure 16: Factor loading heatmaps showing the variable number of factors and the inner variabilities in the influence of the genes on the factors across the three lymphoma subtypes
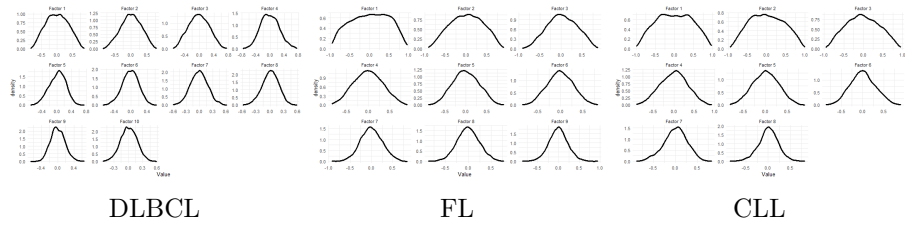
Figure 17: Curves of factor scores showing the distribution patterns of each factors present in a given lymphoma subtype