EM-Driven GMM Optimization for the Randall Dataset: Rooting and Recall

Prabhav Kalaghatgi[1]

[1]Independent Researcher, Hyderabad, India

August 26, 2025

Abstract

1 Introduction

Related work

Claims to clarify

2 Results

2.1 Impact of initial parameters on final log likelihood scores

[width=0.8]"/home/pk/projects/prabhavk.github.io/public/figures/wasm-1756002951415$_{violin_{llfinal_{violin_{t20250824_225734Z.png}}}}$"

Figure 1: This is the caption describing the figure.

2.2 Distribution of optimization scores across root locations

[width=0.8]path/to/your/figure.png

Figure 2: This is the caption describing the figure.

2.3 Wilcoxon-Mann-Whitney tests for difference in log likelihood scores

2.3.1 Impact on initialization method

2.3.2 Impact on root location

row ↓ / col →	h_21	h_22	h_23	h_24	h_25	h_26	h_27	h_28	h_29	h_30	h_31	h_32	h_33	h_34
sample size (per node)	30	30	30	30	30	30	30	30	30	30	30	30	30	30
h_21	—	2	3	3	3	3	3	2	2	3	2	3	3	3
h_22	1	—	3	3	3	3	3	2	2	3	1	3	3	3
h_23	0	0	—	3	3	3	3	1	2	1	0	3	3	3
h_24	0	0	0	—	3	3	3	0	2	0	0	3	3	3
h_25	0	0	0	0	—	3	0	0	0	0	0	3	1	1
h_26	0	0	0	0	0	—	0	0	0	0	0	3	1	1
h_27	0	0	0	0	3	3	—	0	2	0	0	3	3	3
h_28	1	1	2	3	3	3	3	—	3	3	1	3	3	3
h_29	1	1	1	1	3	3	1	0	—	1	0	3	3	3
h_30	0	0	2	3	3	3	3	0	2	—	0	3	3	3
h_31	1	2	3	3	3	3	3	2	3	3	—	3	3	3
h_32	0	0	0	0	0	0	0	0	0	0	0	—	1	0
h_33	0	0	0	0	2	2	0	0	0	0	0	2	—	0
h_34	0	0	0	0	2	2	0	0	0	0	0	3	3	—
h_35	0	0	0	0	2	2	0	0	2	0	0	3	3	3
h_36	0	0	0	2	2	2	2	0	2	0	0	2	2	2
h_37	3	3	3	3	3	3	3	3	3	3	2	3	3	3

Table 1: Agreement counts: number of methods supporting column > row for each node pair.