**Figure 1: Prior Probabilities**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Prior Probability Type** | **P(Gender=0)** | **P(Weight=0|Gender=0)** | **P(Weight=0|Gender=1)** | **P(Height=0|Gender=0)** | **P(Height=0|Gender=1)** |
| G:Provided by Professor | 0.7 | 0.8 | 0.4 | 0.7 | 0.3 |
| T1: Randomly Generated | 0.383271 | 0.511622 | 0.027045 | 0.456521 | 0.96269 |
| T2: Randomly Generated | 0.527978 | 0.816105 | 0.256124 | 0.931043 | 0.141905 |
| T3: Lower extremity | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| T4: Higher extremity | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
| T5: Probability 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |

**Figure 2: Number of Iterations of EM Algorithm for Given case and 5 Test Cases described in Figure 1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 10% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **Iterations** | 5 | 5 | 5 | 6 | 5 | 6 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 30% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **Iterations** | 4 | 5 | 4 | 4 | 6 | 5 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 50% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **Iterations** | 30 | 32 | 29 | 35 | 27 | 33 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 70% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **Iterations** | 65 | 9 | 80 | 82 | 22 | 75 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 100% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **Iterations** | 8 | 20 | 7 | 2 | 2 | 2 |

**Figure 3: Final Probability Gender Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 10% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **P(Gender=0)** | 0.726477 | 0.726481 | 0.726472 | 0.726475 | 0.726481 | 0.726476 |
| **P(Gender=1)** | 0.273523 | 0.273519 | 0.273528 | 0.273525 | 0.273519 | 0.273524 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 30% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **P(Gender=0)** | 0.555546 | 0.555563 | 0.555383 | 0.555503 | 0.555464 | 0.555512 |
| **P(Gender=1)** | 0.444454 | 0.444437 | 0.444617 | 0.444497 | 0.444536 | 0.444488 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 50% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **P(Gender=0)** | 0.704226 | 0.704211 | 0.704219 | 0.704215 | 0.704222 | 0.704217 |
| **P(Gender=1)** | 0.295774 | 0.295789 | 0.295781 | 0.295785 | 0.295778 | 0.295783 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 70% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **P(Gender=0)** | 0.573436 | 0.576313 | 0.573502 | 0.573483 | 0.573507 | 0.573511 |
| **P(Gender=1)** | 0.426564 | 0.423687 | 0.426498 | 0.426517 | 0.426493 | 0.426489 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 100% Missing | **G** | **T1** | **T2** | **T3** | **T4** | **T5** |
| **P(Gender=0)** | 0.720806 | 0.766678 | 0.62245 | 0.1 | 0.9 | 0.5 |
| **P(Gender=1)** | 0.279194 | 0.233322 | 0.37755 | 0.9 | 0.1 | 0.5 |

**Figure 4: Weight|Gender Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **10% Missing** | **P(Weight=0|Gender=0)** | **P(Weight=1|Gender=0)** | **P(Weight=0|Gender=1)** | **P(Weight=1|Gender=1)** |
| **G** | **0.862349** | **0.137651** | **0.634399** | **0.365601** |
| **T1** | **0.86235** | **0.13765** | **0.634394** | **0.365606** |
| **T2** | **0.862348** | **0.137652** | **0.634407** | **0.365593** |
| **T3** | **0.862349** | **0.137651** | **0.634403** | **0.365597** |
| **T4** | **0.86235** | **0.13765** | **0.634395** | **0.365605** |
| **T5** | **0.862349** | **0.137651** | **0.634402** | **0.365598** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **30% Missing** | **P(Weight=0|Gender=0)** | **P(Weight=1|Gender=0)** | **P(Weight=0|Gender=1)** | **P(Weight=1|Gender=1)** |
| **G** | **0.892362** | **0.107638** | **0.234562** | **0.765438** |
| **T1** | **0.892328** | **0.107672** | **0.234579** | **0.765421** |
| **T2** | **0.892667** | **0.107333** | **0.234423** | **0.765577** |
| **T3** | **0.892397** | **0.107603** | **0.234582** | **0.765418** |
| **T4** | **0.892513** | **0.107487** | **0.234494** | **0.765506** |
| **T5** | **0.892423** | **0.107577** | **0.234537** | **0.765463** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **50% Missing** | **P(Weight=0|Gender=0)** | **P(Weight=1|Gender=0)** | **P(Weight=0|Gender=1)** | **P(Weight=1|Gender=1)** |
| **G** | 0.69528 | 0.30472 | 0.373143 | 0.626857 |
| **T1** | 0.695288 | 0.304712 | 0.373139 | 0.626861 |
| **T2** | 0.695284 | 0.304716 | 0.373141 | 0.626859 |
| **T3** | 0.695286 | 0.304714 | 0.37314 | 0.62686 |
| **T4** | 0.695282 | 0.304718 | 0.373142 | 0.626858 |
| **T5** | 0.695285 | 0.304715 | 0.373141 | 0.626859 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **70% Missing** | **P(Weight=0|Gender=0)** | **P(Weight=1|Gender=0)** | **P(Weight=0|Gender=1)** | **P(Weight=1|Gender=1)** |
| **G** | 0.502635 | 0.497365 | 0.027595 | 0.972405 |
| **T1** | 0.504201 | 0.495799 | 0.02224 | 0.97776 |
| **T2** | 0.502683 | 0.497317 | 0.027457 | 0.972543 |
| **T3** | 0.502668 | 0.497332 | 0.027498 | 0.972502 |
| **T4** | 0.502686 | 0.497314 | 0.027448 | 0.972552 |
| **T5** | 0.502689 | 0.497311 | 0.027439 | 0.972561 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **100% Missing** | **P(Weight=0|Gender=0)** | **P(Weight=1|Gender=0)** | **P(Weight=0|Gender=1)** | **P(Weight=1|Gender=1)** |
| **G** | 0.831221 | 0.168779 | 0.361221 | 0.638779 |
| **T1** | 0.823473 | 0.176527 | 0.294277 | 0.705723 |
| **T2** | 0.80374 | 0.19626 | 0.528968 | 0.471032 |
| **T3** | 0.7 | 0.3 | 0.7 | 0.3 |
| **T4** | 0.7 | 0.3 | 0.7 | 0.3 |
| **T5** | 0.7 | 0.3 | 0.7 | 0.3 |

**Figure 5: Height|Gender Table**

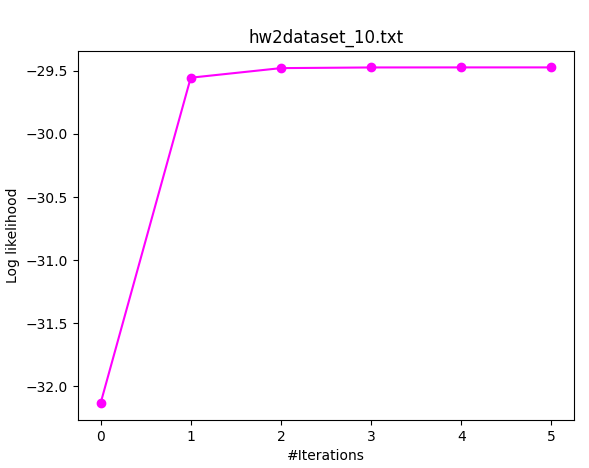
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **10% Missing** | **P(Height=0|Gender=0)** | **P(Height=1|Gender=0)** | **P(Height=0|Gender=1)** | **P(Height=1|Gender=1)** |
| **G** | 0.688252 | 0.311748 | 4.480958e-07 | 1.0 |
| **T1** | 0.688248 | 0.311752 | 7.822644e-07 | 0.999999 |
| **T2** | 0.688258 | 0.311742 | 2.295968e-07 | 1.0 |
| **T3** | 0.688255 | 0.311745 | 4.158310e-07 | 1.0 |
| **T4** | 0.688249 | 0.311751 | 5.168287e-07 | 0.999999 |
| **T5** | 0.688254 | 0.311746 | 2.123324e-07 | 1.0 |

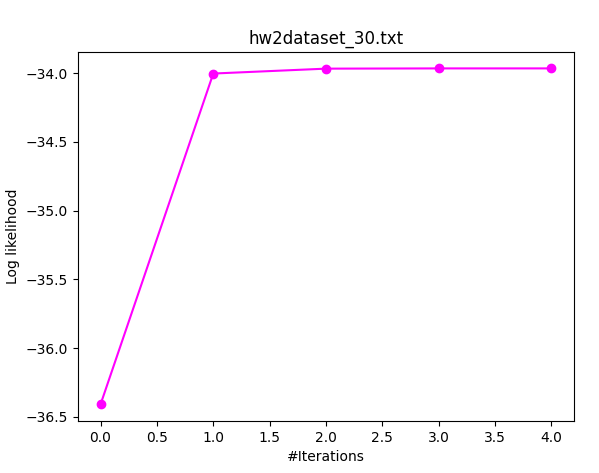
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **30% Missing** | **P(Height=0|Gender=0)** | **P(Height=1|Gender=0)** | **P(Height=0|Gender=1)** | **P(Height=1|Gender=1)** |
| **G** | 0.532355 | 0.467645 | 0.234562 | 0.765438 |
| **T1** | 0.532333 | 0.467667 | 0.234579 | 0.765421 |
| **T2** | 0.532555 | 0.467445 | 0.234423 | 0.765577 |
| **T3** | 0.532363 | 0.467637 | 0.234582 | 0.765418 |
| **T4** | 0.532454 | 0.467546 | 0.234494 | 0.765506 |
| **T5** | 0.532394 | 0.467606 | 0.234537 | 0.765463 |

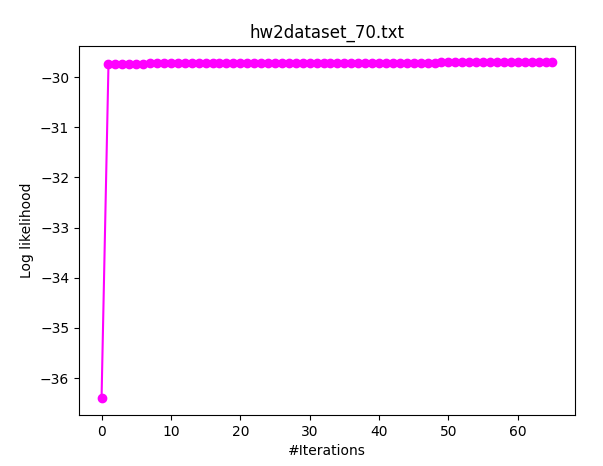
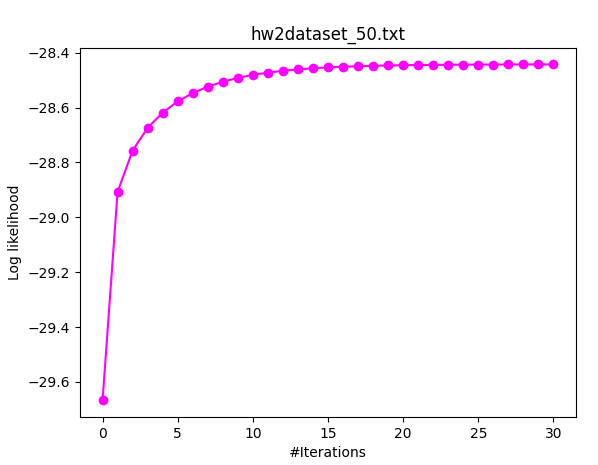
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **50% Missing** | **P(Height=0|Gender=0)** | **P(Height=1|Gender=0)** | **P(Height=0|Gender=1)** | **P(Height=1|Gender=1)** |
| **G** | 0.851907 | 0.148093 | 0.000221 | 0.999779 |
| **T1** | 0.851903 | 0.148097 | 0.00027 | 0.99973 |
| **T2** | 0.851905 | 0.148095 | 0.000243 | 0.999757 |
| **T3** | 0.851904 | 0.148096 | 0.000259 | 0.999741 |
| **T4** | 0.851906 | 0.148094 | 0.000233 | 0.999767 |
| **T5** | 0.851905 | 0.148095 | 0.000249 | 0.999751 |

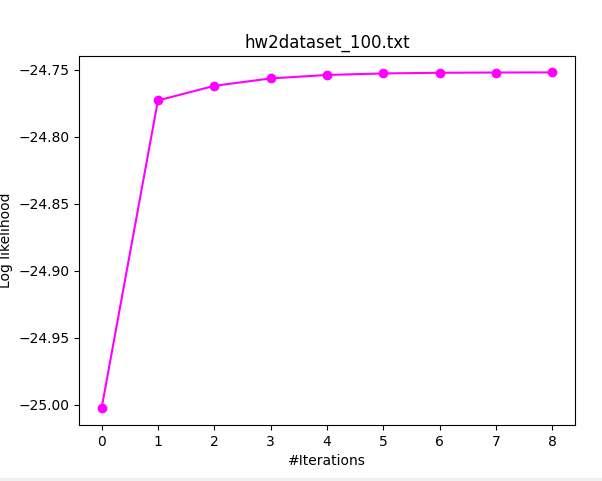
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **70% Missing** | **P(Height=0|Gender=0)** | **P(Height=1|Gender=0)** | **P(Height=0|Gender=1)** | **P(Height=1|Gender=1)** |
| **G** | 0.546578 | 0.453422 | 0.202953 | 0.797047 |
| **T1** | 0.544191 | 0.455809 | 0.203867 | 0.796133 |
| **T2** | 0.546521 | 0.453479 | 0.202977 | 0.797023 |
| **T3** | 0.546538 | 0.453462 | 0.20297 | 0.79703 |
| **T4** | 0.546517 | 0.453483 | 0.202978 | 0.797022 |
| **T5** | 0.546513 | 0.453487 | 0.20298 | 0.79702 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **100% Missing** | **P(Height=0|Gender=0)** | **P(Height=1|Gender=0)** | **P(Height=0|Gender=1)** | **P(Height=1|Gender=1)** |
| **G** | 0.7813 | 0.2187 | 0.311017 | 0.688983 |
| **T1** | 0.760022 | 0.239978 | 0.288478 | 0.711522 |
| **T2** | 0.914255 | 0.085745 | 0.214335 | 0.785665 |
| **T3** | 0.65 | 0.35 | 0.65 | 0.35 |
| **T4** | 0.65 | 0.35 | 0.65 | 0.35 |
| **T5** | 0.65 | 0.35 | 0.65 | 0.35 |

**Fig 6: Graphs for G test(Given by the Professor)**

****

****

****

**Q1: Do multiple starting points help in finding better solutions?**

**A1:** Running EM multiple times with multiple randomized starting points provides the best results. I only used 2 randomized starting points but more would have decreased the number of iterations.

**Q2: Do some of the different solutions have the same log likelihood scores?**

**A2:** Some of the solutions have similar log likelihoods but that could be due to approximation. Various starting points affect the solution. The final probability tables show that the probabilities are similar but differ by a small margin. The final answer is thus NO.

**Q3: How does the data missing rate affect your algorithm and the results?**

**A3:** The data missing rate means that the E-step is called more times and runs for longer. For all of the missing data sets except for 100% missing, the final probabilities differed by a little or a lot(depending on the points used). Obviously, 5 points is not enough because the EM algorithm is designed to be run on multiple randomized starting points. For 100% missing data, different results were produced each time because the Computer must perform many predictions given that no row is complete. In essence, the effectiveness of EM on 100% missing data depends almost exclusively on how well the starting points have been chosen.