Angelica Contero 11/13/13

AI – Project #2 Report Professor Yuan

Starting Parameters Of The Learning:

initial P(G=0) = 0.7

initial P(G=1) = 0.3

initial P(W=0|G=0) = 0.8

initial P(W=0|G=1) = 0.4

initial P(W=1|G=0) = 0.2

initial P(W=1|G=1) = 0.6

initial P(H=0|G=0) = 0.7

initial P(H=0|G=1) = 0.3

initial P(H=1|G=0) = 0.3

initial P(H=1|G=1) = 0.7

convergence at < 0.0001

hw2dataset\_10 Convergence Table:

|  |  |
| --- | --- |
| Iteration # | Comparison Difference |
| 1 | 100 (starting default) |
| 2 | 5.9844723 |
| 3 | 0.11460304 |
| 4 | 0.008042794 |
| 5 | 5.7561696E-4 |
| 6 | 4.2829663E-5 |

Last used probability tables for the learning:

Newly calculated values for next iteration:

new P(G=0) = 0.6906436

new P(G=1) = 0.30935666

new P(W=0|G=0) = 0.82942766

new P(W=0|G=1) = 0.42721435

new P(W=1|G=0) = 0.17057228

new P(W=1|G=1) = 0.5727867

new P(H=0|G=0) = 0.6574697

new P(H=0|G=1) = 0.31007242

new P(H=1|G=0) = 0.34252995

new P(H=1|G=1) = 0.6899281

**NOTE:** For all probability tables listed in this report, newer tables were created right after the one I’m displaying which were to be used for the next run, but since the program had reached convergence during the current run, the newer tables were never utilized. These were the last **used** ones.

hw2dataset\_30 Convergence Table:

|  |  |
| --- | --- |
| Iteration # | Comparison Difference |
| 1 | 100 (starting default) |
| 2 | 5.9412026 |
| 3 | 1.256608 |
| 4 | 0.3671767 |
| 5 | 0.13155618 |
| 6 | 0.0494712 |
| 7 | 0.01886775 |
| 8 | 0.007239476 |
| 9 | 0.0027911402 |
| 10 | 0.0010905601 |
| 11 | 4.3053925E-4 |
| 12 | 1.6806275E-4 |
| 13 | 5.563721E-5 |

Last used probability tables for the learning:

Newly calculated values for next iteration:

new P(G=0) = 0.6815607

new P(G=1) = 0.31843913

new P(W=0|G=0) = 0.7858988

new P(W=0|G=1) = 0.3748352

new P(W=1|G=0) = 0.21410143

new P(W=1|G=1) = 0.6251646

new P(H=0|G=0) = 0.6509015

new P(H=0|G=1) = 0.3497404

new P(H=1|G=0) = 0.34909865

new P(H=1|G=1) = 0.65025955

hw2dataset\_50 Convergence Table:

|  |  |
| --- | --- |
| Iteration # | Comparison Difference |
| 1 | 100 (starting default) |
| 2 | 7.8160915 |
| 3 | 3.2712584 |
| 4 | 1.5220432 |
| 5 | 0.7255813 |
| 6 | 0.35368305 |
| 7 | 0.17655468 |
| 8 | 0.09051508 |
| 9 | 0.04785471 |
| 10 | 0.026102781 |
| 11 | 0.0147008 |
| 12 | 0.009197693 |
| 13 | 0.0061631687 |
| 14 | 0.004189484 |
| 15 | 0.0028428435 |
| 16 | 0.0019454248 |
| 17 | 0.0013153665 |
| 18 | 9.058751E-4 |
| 19 | 6.291382E-4 |
| 20 | 4.3012574E-4 |
| 21 | 2.8929114E-4 |
| 22 | 2.0503253E-4 |
| 23 | 1.293309E-4 |
| 24 | 9.059906E-5 |

Last used probability tables for the learning:

Newly calculated values for next iteration:

new P(G=0) = 0.6627312

new P(G=1) = 0.33726865

new P(W=0|G=0) = 0.7441548

new P(W=0|G=1) = 0.47981134

new P(W=1|G=0) = 0.25584534

new P(W=1|G=1) = 0.5201887

new P(H=0|G=0) = 0.68823826

new P(H=0|G=1) = 0.33766267

new P(H=1|G=0) = 0.3117621

new P(H=1|G=1) = 0.6623371

hw2dataset\_70 Convergence Table:

|  |  |
| --- | --- |
| Iteration # | Comparison Difference |
| 1 | 100 (starting default) |
| 2 | 2.7471147 |
| 3 | 1.2421175 |
| 4 | 0.70462155 |
| 5 | 0.39946473 |
| 6 | 0.22690816 |
| 7 | 0.12939289 |
| 8 | 0.0742222 |
| 9 | 0.042927325 |
| 10 | 0.02508558 |
| 11 | 0.015257895 |
| 12 | 0.009460356 |
| 13 | 0.0060943775 |
| 14 | 0.004309293 |
| 15 | 0.0031920709 |
| 16 | 0.0022932515 |
| 17 | 0.0017804019 |
| 18 | 0.0012729168 |
| 19 | 9.483583E-4 |
| 20 | 6.6273287E-4 |
| 21 | 5.069338E-4 |
| 22 | 3.3191964E-4 |
| 23 | 2.894327E-4 |
| 24 | 1.5406683E-4 |
| 25 | 1.12742186E-4 |
| 26 | 9.2320144E-5 |

Last used probability tables for the learning:

Newly calculated values for next iteration:

new P(G=0) = 0.6906436

new P(G=1) = 0.30935666

new P(W=0|G=0) = 0.82942766

new P(W=0|G=1) = 0.42721435

new P(W=1|G=0) = 0.17057228

new P(W=1|G=1) = 0.5727867

new P(H=0|G=0) = 0.6574697

new P(H=0|G=1) = 0.31007242

new P(H=1|G=0) = 0.34252995

new P(H=1|G=1) = 0.6899281

hw2dataset\_100 Convergence Table:

|  |  |
| --- | --- |
| Iteration # | Comparison Difference |
| 1 | 100 (starting default) |
| 2 | 1.1589036 |
| 3 | 0.8369533 |
| 4 | 0.59684443 |
| 5 | 0.4215872 |
| 6 | 0.2958457 |
| 7 | 0.20658341 |
| 8 | 0.14377102 |
| 9 | 0.09983596 |
| 10 | 0.06933703 |
| 11 | 0.047958456 |
| 12 | 0.03309615 |
| 13 | 0.022871561 |
| 14 | 0.015863545 |
| 15 | 0.011043839 |
| 16 | 0.0075542405 |
| 17 | 0.005201496 |
| 18 | 0.0036379397 |
| 19 | 0.00250528 |
| 20 | 0.0017594844 |
| 21 | 0.0012085885 |
| 22 | 7.74242E-4 |
| 23 | 5.8440864E-4 |
| 24 | 4.0411204E-4 |
| 25 | 2.7930737E-4 |
| 26 | 1.8775463E-4 |
| 27 | 1.2767315E-4 |
| 28 | 1.2235343E-4 |
| 29 | 7.727742E-5 |

Last used probability tables for the learning:

Newly calculated values for next iteration:

new P(G=0) = 0.69669783

new P(G=1) = 0.30330238

new P(W=0|G=0) = 0.7966559

new P(W=0|G=1) = 0.36258262

new P(W=1|G=0) = 0.20334484

new P(W=1|G=1) = 0.63741714

new P(H=0|G=0) = 0.7250349

new P(H=0|G=1) = 0.29630494

new P(H=1|G=0) = 0.27496523

new P(H=1|G=1) = 0.7036948

SUMMARY OF ALL LEARNING TESTS WITH ALL THE DIFFERENT SETS OF STARTING PARAMETERS:

TEST 1: Using the given starting parameters.

|  |  |  |
| --- | --- | --- |
| Dataset | #Iterations | Final Comparison Difference |
| hw2dataset\_10 | 6 | 4.2829663E-5 |
| hw2dataset\_30 | 13 | 5.563721E-5 |
| hw2dataset\_50 | 24 | 9.059906E-5 |
| hw2dataset\_70 | 26 | 9.2320144E-5 |
| hw2dataset\_100 | 29 | 7.727742E-5 |

TEST 2: Using the starting parameters shown below.

|  |  |  |
| --- | --- | --- |
| Dataset | #Iterations | Final Comparison Difference |
| hw2dataset\_10 | 6 | 5.953014E-5 |
| hw2dataset\_30 | 14 | 5.6859106E-5 |
| hw2dataset\_50 | 31 | 6.0111284E-5 |
| hw2dataset\_70 | 27 | 9.915605E-5 |
| hw2dataset\_100 | 28 | 6.553903E-5 |

TEST 2 Parameters:

//increased larger one by .1 and decreased smaller one by .1

*Gprob*[0] = 0.8f;

*Gprob*[1] = 0.2f;

*WgivenGprob*[0][0] = 0.9f; // W|G

*WgivenGprob*[0][1] = 0.5f; // W|~G

*WgivenGprob*[1][0] = 0.1f; // ~W|G

*WgivenGprob*[1][1] = 0.5f; //~W|~G

*HgivenGprob*[0][0] = 0.8f; // H|G

*HgivenGprob*[0][1] = 0.2f; // H|~G

*HgivenGprob*[1][0] = 0.2f; // ~H|G

*HgivenGprob*[1][1] = 0.8f; //~H|~G

TEST 3: Using the starting parameters shown below.

|  |  |  |
| --- | --- | --- |
| Dataset | #Iterations | Final Comparison Difference |
| hw2dataset\_10 | 7 | 2.099201E-5 |
| hw2dataset\_30 | 12 | 5.384907E-5 |
| hw2dataset\_50 | 31 | 6.299466E-5 |
| hw2dataset\_70 | 37 | 9.519234E-5 |
| hw2dataset\_100 | 41 | 9.843707E-5 |

TEST 3 Parameters:

//decreased larger one by .1 and increased smaller one by .1

*Gprob*[0] = 0.6f;

*Gprob*[1] = 0.4f;

*WgivenGprob*[0][0] = 0.7f; // W|G

*WgivenGprob*[0][1] = 0.5f; // W|~G

*WgivenGprob*[1][0] = 0.3f; // ~W|G

*WgivenGprob*[1][1] = 0.5f; //~W|~G

*HgivenGprob*[0][0] = 0.6f; // H|G

*HgivenGprob*[0][1] = 0.4f; // H|~G

*HgivenGprob*[1][0] = 0.4f; // ~H|G

*HgivenGprob*[1][1] = 0.6f; //~H|~G

TEST 4: Using the starting parameters shown below.

|  |  |  |
| --- | --- | --- |
| Dataset | #Iterations | Final Comparison Difference |
| hw2dataset\_10 | 7 | 8.453801E-5 |
| hw2dataset\_30 | 15 | 4.61936E-5 |
| hw2dataset\_50 | 34 | 9.064749E-5 |
| hw2dataset\_70 | 42 | 9.519234E-5 |
| hw2dataset\_100 | 30 | 9.8243356E-5 |

TEST 4 Parameters:

//subtracted .5 from larger number, added .5 to the smaller number

*Gprob*[0] = 0.2f;

*Gprob*[1] = 0.8f;

*WgivenGprob*[0][0] = 0.3f; // W|G

*WgivenGprob*[0][1] = 0.9f; // W|~G

*WgivenGprob*[1][0] = 0.7f; // ~W|G

*WgivenGprob*[1][1] = 0.1f; //~W|~G

*HgivenGprob*[0][0] = 0.2f; // H|G

*HgivenGprob*[0][1] = 0.8f; // H|~G

*HgivenGprob*[1][0] = 0.8f; // ~H|G

*HgivenGprob*[1][1] = 0.2f; //~H|~G

TEST 5: Using the starting parameters shown below.

|  |  |  |
| --- | --- | --- |
| Dataset | #Iterations | Final Comparison Difference |
| hw2dataset\_10 | 7 | 2.9325485E-5 |
| hw2dataset\_30 | 14 | 5.9057027E-5 |
| hw2dataset\_50 | 33 | 6.3069165E-5 |
| hw2dataset\_70 | 40 | 9.1489404E-5 |
| hw2dataset\_100 | 2 | 0.0 |

TEST 5 Parameters:

//everything is .5

*Gprob*[0] = 0.5f;

*Gprob*[1] = 0.5f;

*WgivenGprob*[0][0] = 0.5f; // W|G

*WgivenGprob*[0][1] = 0.5f; // W|~G

*WgivenGprob*[1][0] = 0.5f; // ~W|G

*WgivenGprob*[1][1] = 0.5f; //~W|~G

*HgivenGprob*[0][0] = 0.5f; // H|G

*HgivenGprob*[0][1] = 0.5f; // H|~G

*HgivenGprob*[1][0] = 0.5f; // ~H|G

*HgivenGprob*[1][1] = 0.5f; //~H|~G

Analysis:

1. Do multiple starting points help in finding better solutions?
   * Based on my results, I would say that using multiple starting points is not absolutely necessary. The #of iterations remains similar across all sets of starting points which means to me that as long as you have a strict convergence measure, you will find an pretty accurate data set in a similar amount of time, no matter which starting parameters you use.
   * Nevertheless, the varied starting parameters create a discrepancy among all the tests by at most 0.03 (in examining the updated probabilities tables of P(G),P(W|G) and P(H|G). If you would like to get more accurate than that, then perhaps comparison with multiple starting points would be helpful and then you could take the average or take the most frequent values, something to that effect.
   * With regards to the hw2dataset\_100, I noticed an interesting trend. It would seem that if you choose “ideal” parameters for this data set, the algorithm will take very few iterations to create a data set that meets convergence. However, given a completely empty G column, one would assume that the probability of G is 50/50 since it’s either 0 or 1. I think that this is why when I choose probabilities of .5, it finds a solution in 2 iterations. However, I cannot speak to the full accuracy of the data in this table since I don’t have a complete data set to compare it to. That being said, its values are still within 0.03 of the values of all the other tests. If this is, in fact, sufficient, then I would say that my previous conclusion is still true. Multiple starting points is not necessary.
2. Do some of the different solutions have the same likelihood scores?

Analysis of my tables:

Since I was not able to calculate the likelihood scores, I will instead compare final comparison difference values (difference between iterations) listed in my tables. My comparison differences for the last iterations with the original given initial parameters (Test #1), vary between 4.28E-5 (for the hw2dataset\_10) and 9.23E-5 (for the hw2dataset\_100).

Within a test using the same parameters (Tests #2-#5), for all the hw2dataset files, the variations between the files is similar to those of the original parameters test (Test #1). Comparison differences can range from around 4E-5 to around 9E-5 or 10E-5. However for Test #2 and #5 it ranges from around 2E-5 to around 10E-5 and around 3E-5 to around 9E-5, respectively.

Across the varying starting parameters (Tests #1-#5), but between the same hw2dataset files, here are the ranges I get:

* + 1. hw2dataset\_10: 2.09E-5 to 8.45E-5
    2. hw2dataset\_30: 4.61E-5 to 5.9E-5
    3. hw2dataset\_50: 6.011E-5 to 9.06E-5
    4. hw2dataset\_70: 9.9E-5 to 9.14E-5 (the tightest range seen here)
    5. hw2dataset\_100: 0.0 to 9.84E-5 (0.0 is TEST #5!)

**ANSWER:** In conclusion, the closest comparison difference scores are between all of the Tests #1-#5 (varying the parameters) for hw2dataset\_70 which was 9.9E-5 to 9.14E-5. So given any starting parameters, it seems that this dataset will converge with a similar comparison difference between iterations.

1. How does the data missing rate affect your algorithm and the results?

- The greater the percentage of missing data points, the more iterations is needed before convergence, and the longer it takes to run the program (which is still relatively quick).