CS381
Project #2 Report
Jinwei Lin
10/30/2015

Each dataset have five random starting points and the starting parameters given in the
project speciation tested at last.

Data set with missing rate being 10%

The starting points of the learning:
P(gender=M) =0.82
P(w=0|g=0) = 0.37
P(w=0|g=1) = 0.88
P(h=0|g=0) = 0.83
P(h=0|g=1) = 0.77

The final conditional probability tables for hw2dataset_10.txt is:
Gender table
P(gender=M) = 0.642462 P(gender=F) = 0.357538

Weight(w) given gender(g) table
P(w=0|g=0) = 0.794435 P(w=1|g=0) = 0.205565
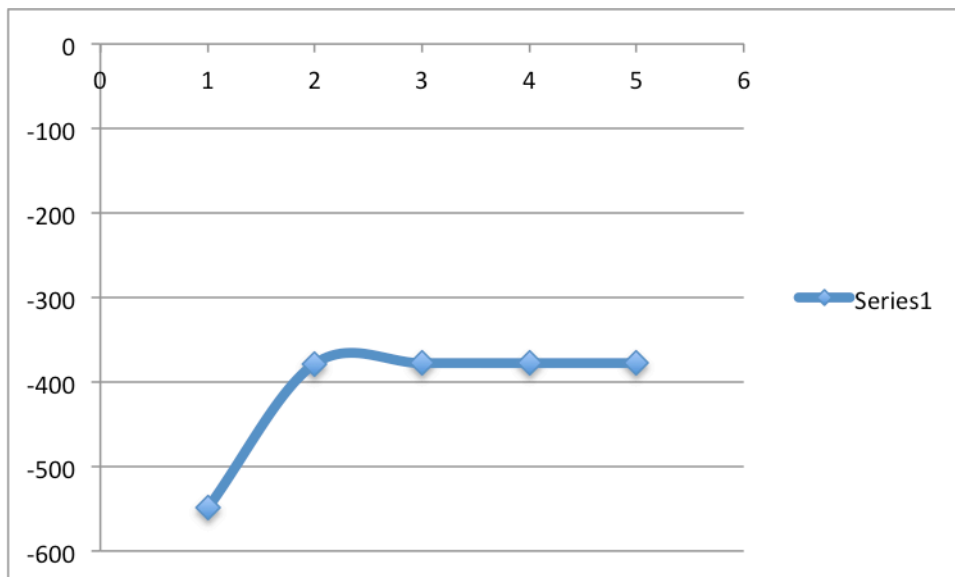P(w=0|g=1) = 0.348511 P(w=1|g=1) = 0.651489

Height(h) given gender(g) table
P(h=0|g=0) = 0.661922 P(h=1|g=0) = 0.338078
P(h=0|g=1) = 0.26498 P(h=1|g=0) = 0.73502

Plots of the likelihood vs number of iterations

The starting points of the learning:
P(gender=M) =0.52
P(w=0|g=0) = 0.11
P(w=0|g=1) = 0.45
P(h=0|g=0) = 0.51
P(h=0|g=1) = 0.13

The final conditional probability tables for hw2dataset_10.txt is:
Gender table
P(gender=M) = 0.642461 P(gender=F) = 0.357539

Weight(w) given gender(g) table
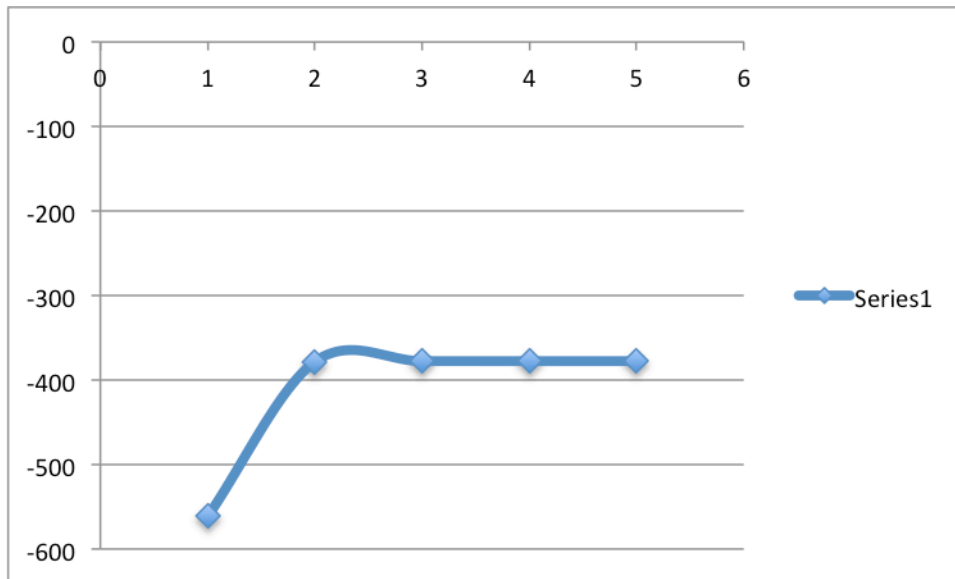P(w=0|g=0) = 0.794435 P(w=1|g=0) = 0.205565
P(w=0|g=1) = 0.348511 P(w=1|g=1) = 0.651489

Height(h) given gender(g) table
P(h=0|g=0) = 0.661922 P(h=1|g=0) = 0.338078
P(h=0|g=1) = 0.264979 P(h=1|g=0) = 0.735021

The starting points of the learning:

P(gender=M) =0.49
P(w=0|g=0) = 0.56
P(w=0|g=1) = 0.65
P(h=0|g=0) = 0.56
P(h=0|g=1) = 0.66

The final conditional probability tables for hw2dataset_10.txt is:
Gender table

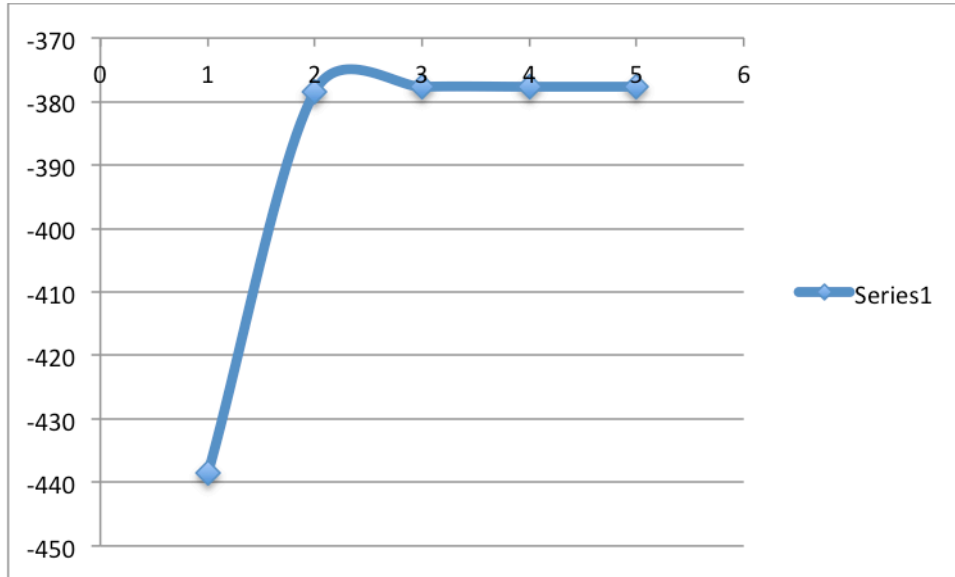P(gender=M) = 0.642462 P(gender=F) = 0.357538

Weight(w) given gender(g) table
P(w=0|g=0) = 0.794435 P(w=1|g=0) = 0.205565
P(w=0|g=1) = 0.348511 P(w=1|g=1) = 0.651489

Height(h) given gender(g) table
P(h=0|g=0) = 0.661922 P(h=1|g=0) = 0.338078
P(h=0|g=1) = 0.26498 P(h=1|g=0) = 0.73502

The starting points of the learning:
P(gender=M) =0.96
P(w=0|g=0) = 0.33
P(w=0|g=1) = 0.21
P(h=0|g=0) = 0.89
P(h=0|g=1) = 0.41

The final conditional probability tables for hw2dataset_10.txt is:
Gender table
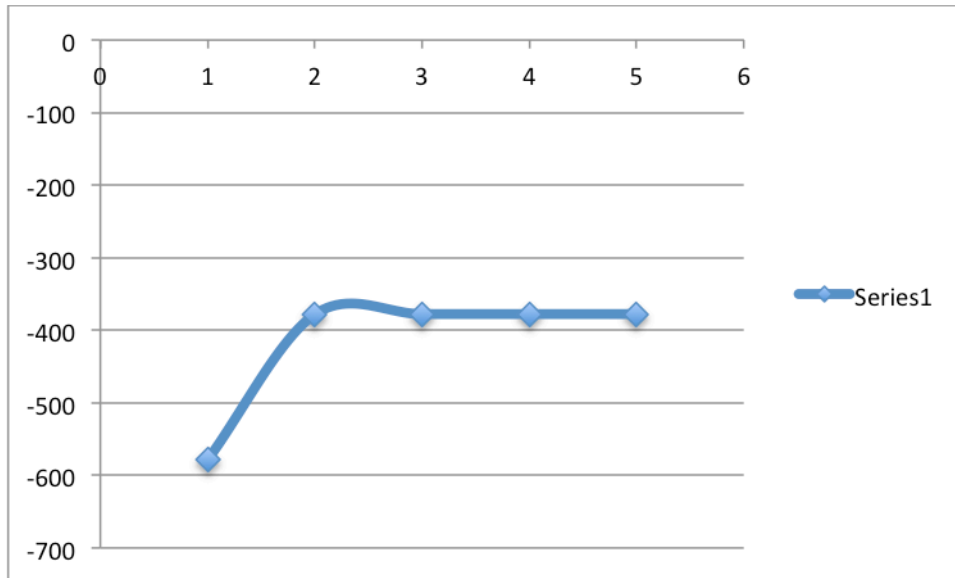P(gender=M) = 0.642463 P(gender=F) = 0.357537

Weight(w) given gender(g) table
P(w=0|g=0) = 0.794435 P(w=1|g=0) = 0.205565
P(w=0|g=1) = 0.34851 P(w=1|g=1) = 0.65149

Height(h) given gender(g) table
P(h=0|g=0) = 0.661922 P(h=1|g=0) = 0.338078
P(h=0|g=1) = 0.26498 P(h=1|g=0) = 0.73502

The starting points of the learning:
P(gender=M) =0.26
P(w=0|g=0) = 0.79
P(w=0|g=1) = 0.95
P(h=0|g=0) = 0.35
P(h=0|g=1) = 0.85

The final conditional probability tables for hw2dataset_10.txt is:
Gender table
P(gender=M) = 0.642462 P(gender=F) = 0.357538

Weight(w) given gender(g) table
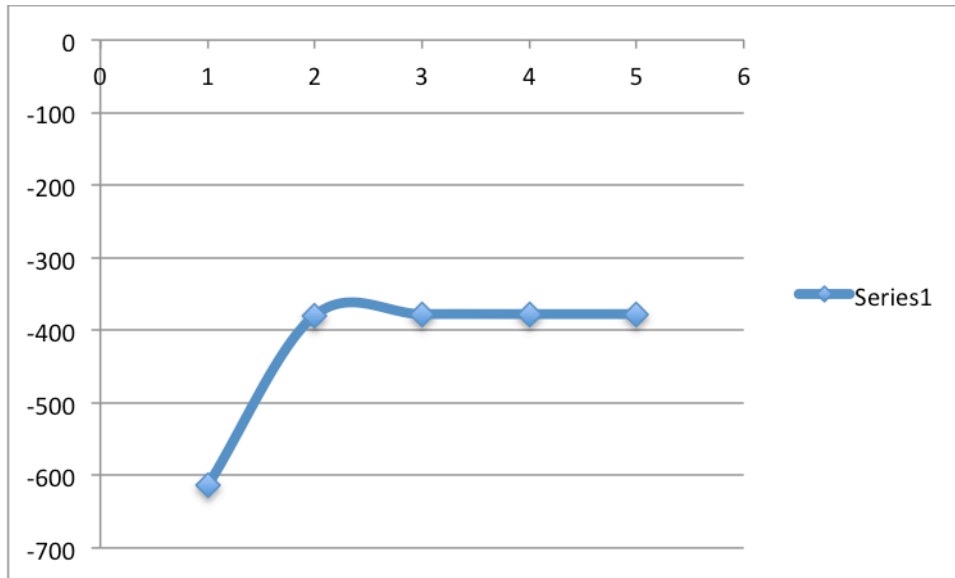P(w=0|g=0) = 0.794435 P(w=1|g=0) = 0.205565
P(w=0|g=1) = 0.348512 P(w=1|g=1) = 0.651488

Height(h) given gender(g) table
P(h=0|g=0) = 0.661922 P(h=1|g=0) = 0.338078
P(h=0|g=1) = 0.26498 P(h=1|g=0) = 0.73502

The final conditional probability tables for hw2dataset_10.txt is:
Gender table
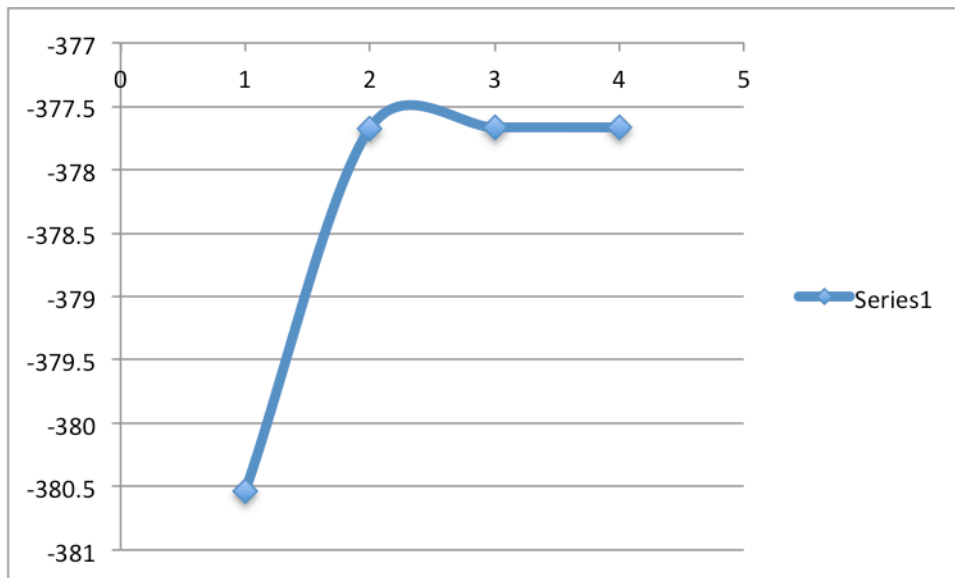P(gender=M) = 0.642462 P(gender=F) = 0.357538

Weight(w) given gender(g) table
P(w=0|g=0) = 0.794435 P(w=1|g=0) = 0.205565
P(w=0|g=1) = 0.34851 P(w=1|g=1) = 0.65149

Height(h) given gender(g) table
P(h=0|g=0) = 0.661922 P(h=1|g=0) = 0.338078
P(h=0|g=1) = 0.264979 P(h=1|g=0) = 0.735021

Data set with missing rate being 30%

The starting points of the learning:
P(gender=M) =0.47
P(w=0|g=0) = 0.35
P(w=0|g=1) = 0.79
P(h=0|g=0) = 0.14
P(h=0|g=1) = 0.17
The final conditional probability tables for hw2dataset_30.txt is:
Gender table
P(gender=M) = 0.681544 P(gender=F) = 0.318456

Weight(w) given gender(g) table
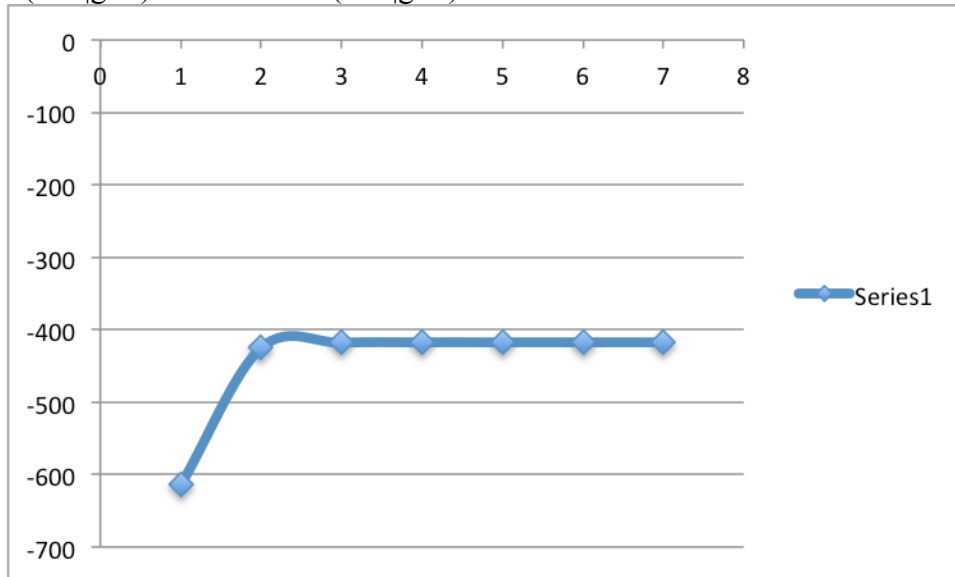P(w=0|g=0) = 0.785886 P(w=1|g=0) = 0.214114
P(w=0|g=1) = 0.374883 P(w=1|g=1) = 0.625117

Height(h) given gender(g) table
P(h=0|g=0) = 0.650897 P(h=1|g=0) = 0.349103
P(h=0|g=1) = 0.349767 P(h=1|g=0) = 0.650233

The starting points of the learning:
P(gender=M) =0.29
P(w=0|g=0) = 0.1
P(w=0|g=1) = 0.12
P(h=0|g=0) = 0.43
P(h=0|g=1) = 0.48
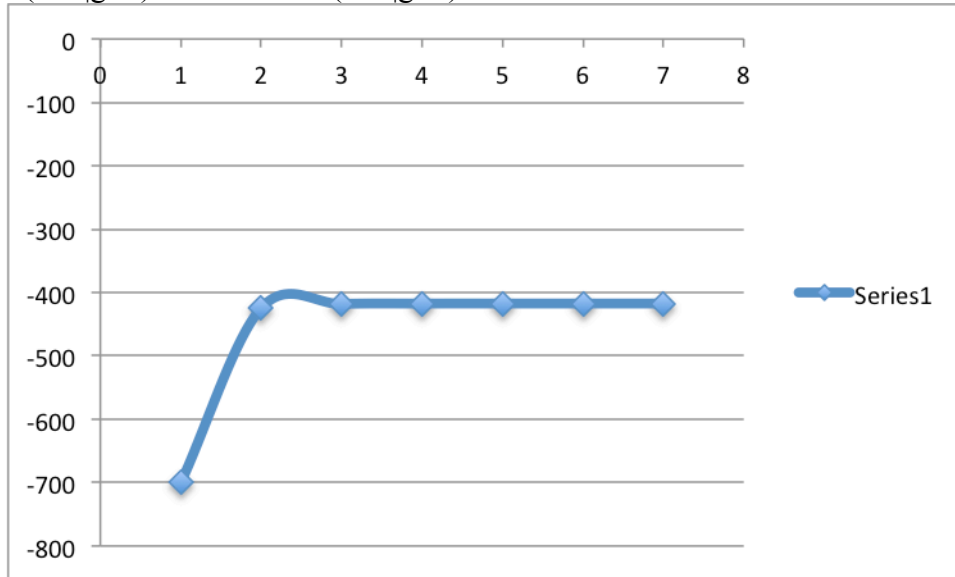The final conditional probability tables for hw2dataset_30.txt is:
Gender table
P(gender=M) = 0.681394 P(gender=F) = 0.318606

Weight(w) given gender(g) table

P(w=0|g=0) = 0.786044 P(w=1|g=0) = 0.213956
P(w=0|g=1) = 0.37474 P(w=1|g=1) = 0.62526

Height(h) given gender(g) table
P(h=0|g=0) = 0.650854 P(h=1|g=0) = 0.349146
P(h=0|g=1) = 0.349999 P(h=1|g=0) = 0.650001

The starting points of the learning:
P(gender=M) =0.31
P(w=0|g=0) = 0.07
P(w=0|g=1) = 0.35
P(h=0|g=0) = 0.51
P(h=0|g=1) = 0.66
The final conditional probability tables for hw2dataset_30.txt is:
Gender table
P(gender=M) = 0.681407 P(gender=F) = 0.318593

Weight(w) given gender(g) table
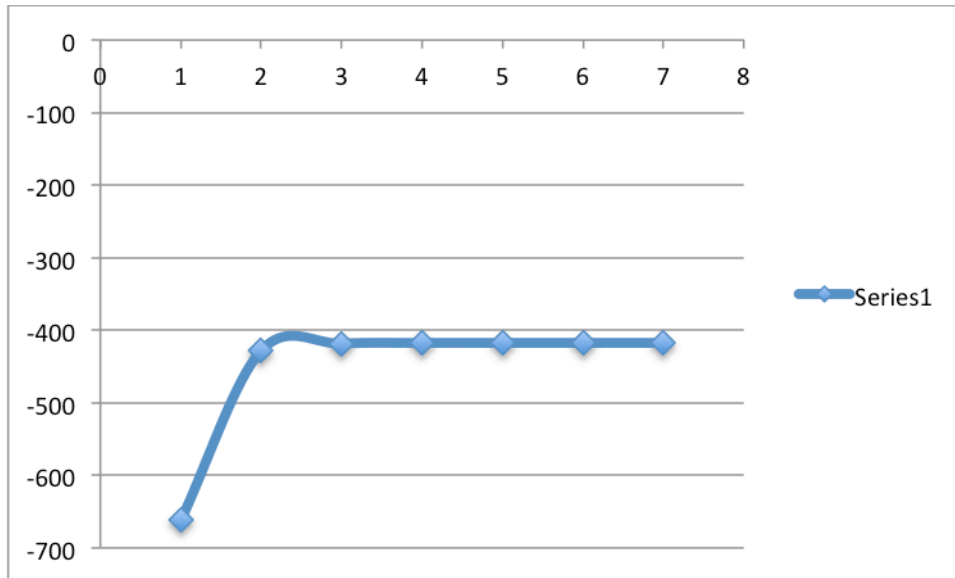P(w=0|g=0) = 0.786022 P(w=1|g=0) = 0.213978
P(w=0|g=1) = 0.374772 P(w=1|g=1) = 0.625228

Height(h) given gender(g) table
P(h=0|g=0) = 0.650858 P(h=1|g=0) = 0.349142
P(h=0|g=1) = 0.349979 P(h=1|g=0) = 0.650021

The starting points of the learning:
P(gender=M) =0
P(w=0|g=0) = 0.67
P(w=0|g=1) = 0.29
P(h=0|g=0) = 0.8
P(h=0|g=1) = 0.84

The final conditional probability tables for hw2dataset_30.txt is:
Gender table
P(gender=M) = 0.70865 P(gender=F) = 0.29135

Weight(w) given gender(g) table
P(w=0|g=0) = 0.718843 P(w=1|g=0) = 0.281157
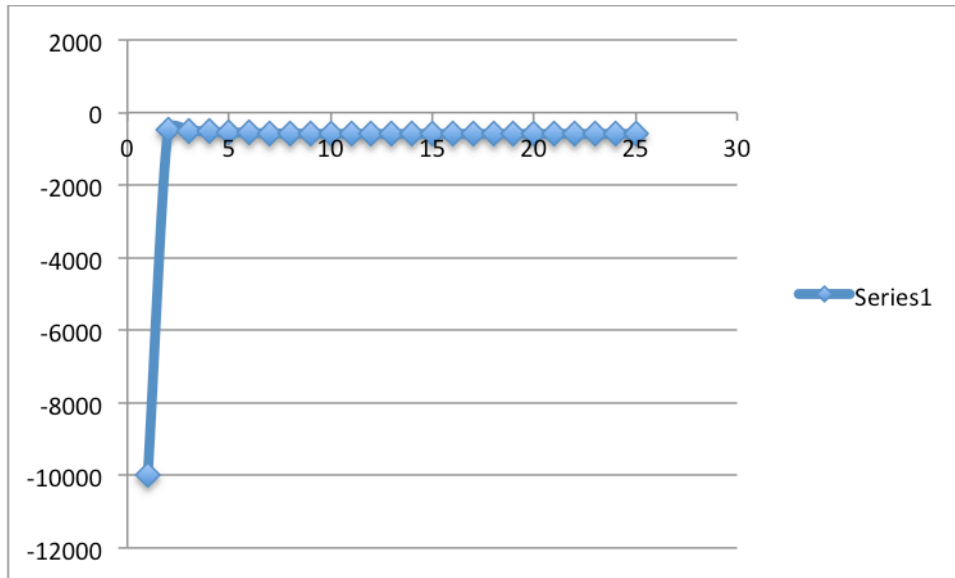P(w=0|g=1) = 0.892396 P(w=1|g=1) = 0.107604

Height(h) given gender(g) table
P(h=0|g=0) = 0.600832 P(h=1|g=0) = 0.399168
P(h=0|g=1) = 0.858073 P(h=1|g=0) = 0.141927

The starting points of the learning:
P(gender=M) =0.96
P(w=0|g=0) = 0.29
P(w=0|g=1) = 0.89
P(h=0|g=0) = 0.16
P(h=0|g=1) = 0.37
The final conditional probability tables for hw2dataset_30.txt is:
Gender table
P(gender=M) = 0.681712 P(gender=F) = 0.318288

Weight(w) given gender(g) table
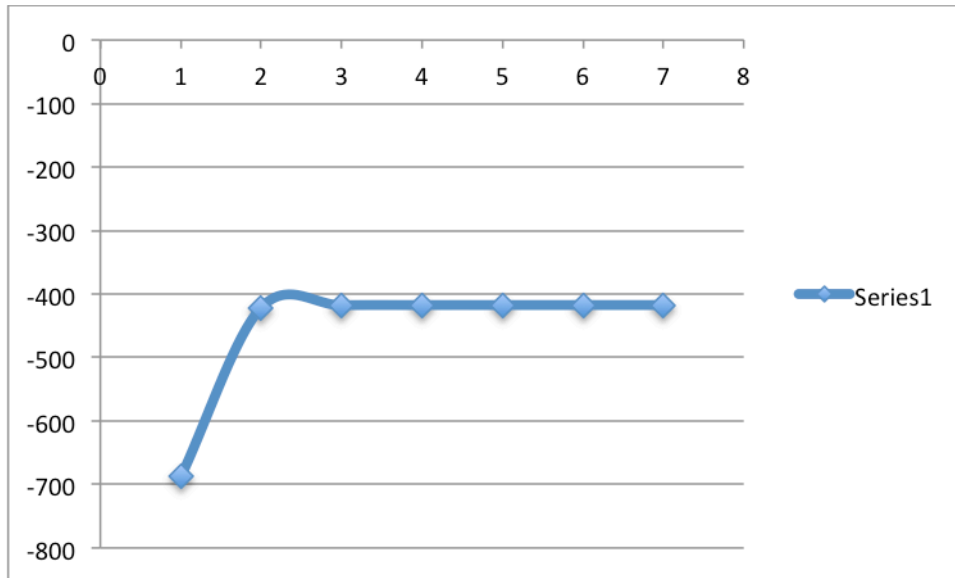P(w=0|g=0) = 0.785739 P(w=1|g=0) = 0.214261
P(w=0|g=1) = 0.374982 P(w=1|g=1) = 0.625018

Height(h) given gender(g) table
P(h=0|g=0) = 0.65094 P(h=1|g=0) = 0.34906
P(h=0|g=1) = 0.349515 P(h=1|g=0) = 0.650485

The final conditional probability tables for hw2dataset_30.txt is:
Gender table
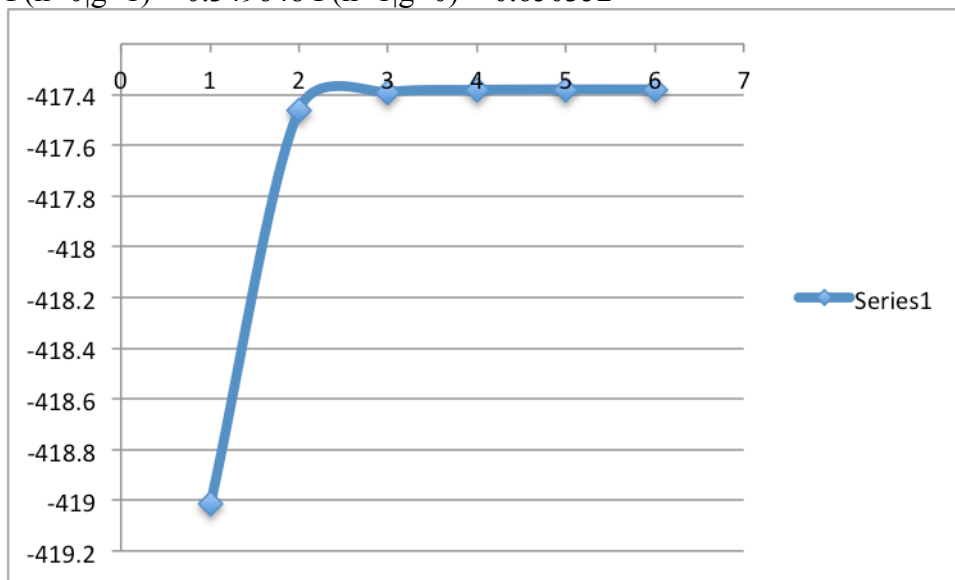P(gender=M) = 0.681612 P(gender=F) = 0.318388

Weight(w) given gender(g) table
P(w=0|g=0) = 0.785845 P(w=1|g=0) = 0.214155
P(w=0|g=1) = 0.374884 P(w=1|g=1) = 0.625116

Height(h) given gender(g) table
P(h=0|g=0) = 0.650922 P(h=1|g=0) = 0.349078
P(h=0|g=1) = 0.349648 P(h=1|g=0) = 0.650352

Data set with missing rate being 50%

The starting points of the learning:
P(gender=M) =0.85
P(w=0|g=0) = 0.42
P(w=0|g=1) = 0.23
P(h=0|g=0) = 0.94
P(h=0|g=1) = 0.6
The final conditional probability tables for hw2dataset_50.txt is:
Gender table
P(gender=M) = 0.663596 P(gender=F) = 0.336404

Weight(w) given gender(g) table
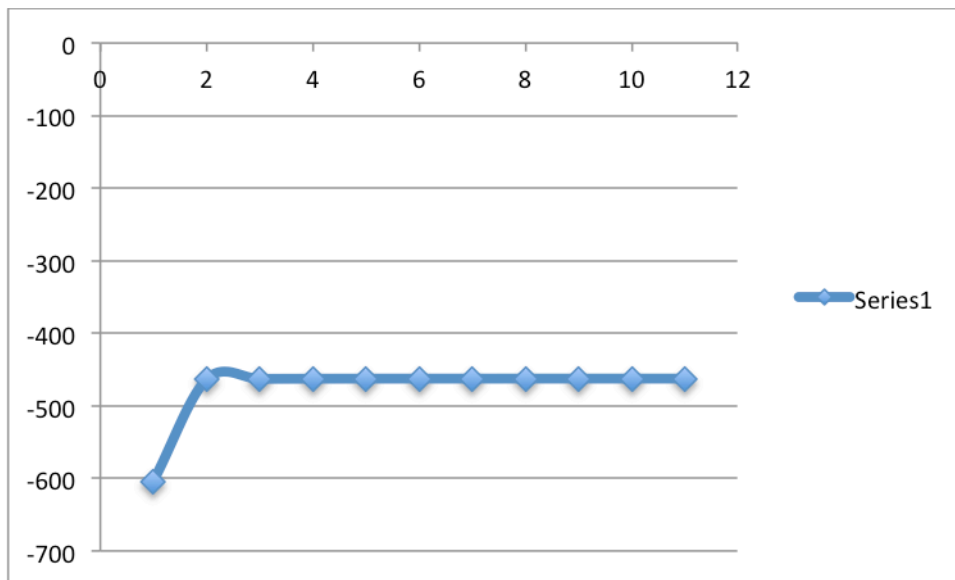P(w=0|g=0) = 0.743306 P(w=1|g=0) = 0.256694
P(w=0|g=1) = 0.480807 P(w=1|g=1) = 0.519193

Height(h) given gender(g) table
P(h=0|g=0) = 0.688703 P(h=1|g=0) = 0.311297
P(h=0|g=1) = 0.335845 P(h=1|g=0) = 0.664155

The starting points of the learning:
P(gender=M) =0.99
P(w=0|g=0) = 0.29
P(w=0|g=1) = 0.03
P(h=0|g=0) = 0.01
P(h=0|g=1) = 0.35
The final conditional probability tables for hw2dataset_50.txt is:
Gender table
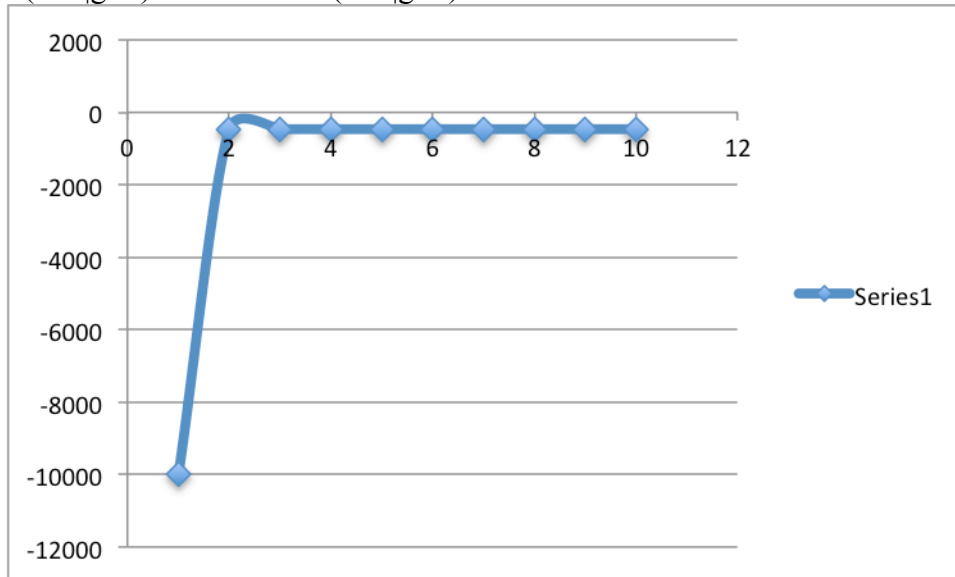P(gender=M) = 0.663445 P(gender=F) = 0.336555

Weight(w) given gender(g) table
P(w=0|g=0) = 0.743666 P(w=1|g=0) = 0.256334
P(w=0|g=1) = 0.480214 P(w=1|g=1) = 0.519786

Height(h) given gender(g) table
P(h=0|g=0) = 0.688383 P(h=1|g=0) = 0.311617
P(h=0|g=1) = 0.336634 P(h=1|g=0) = 0.663366

The starting points of the learning:
P(gender=M) =0.55
P(w=0|g=0) = 0.63
P(w=0|g=1) = 0.8
P(h=0|g=0) = 0.15
P(h=0|g=1) = 0.96
The final conditional probability tables for hw2dataset_50.txt is:
Gender table
P(gender=M) = 0.662266 P(gender=F) = 0.337734

Weight(w) given gender(g) table
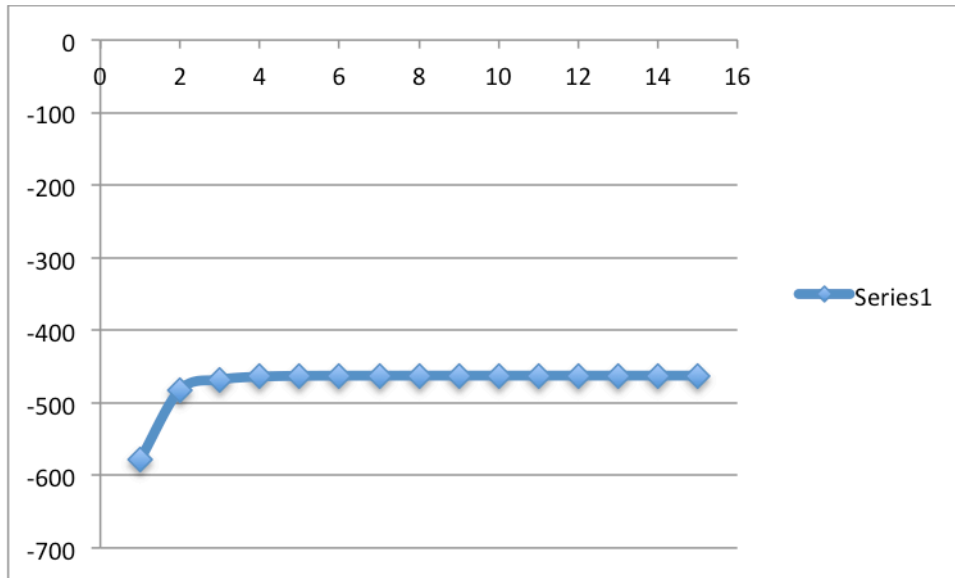P(w=0|g=0) = 0.744622 P(w=1|g=0) = 0.255378
P(w=0|g=1) = 0.479259 P(w=1|g=1) = 0.520741

Height(h) given gender(g) table
P(h=0|g=0) = 0.68798 P(h=1|g=0) = 0.31202
P(h=0|g=1) = 0.338651 P(h=1|g=0) = 0.661349

The starting points of the learning:

P(gender=M) =0.47

P(w=0|g=0) = 0.2

P(w=0|g=1) = 0.22

P(h=0|g=0) = 0.26

P(h=0|g=1) = 0.75

The final conditional probability tables for hw2dataset_50.txt is:

Gender table

P(gender=M) = 0.66226 P(gender=F) = 0.33774

Weight(w) given gender(g) table
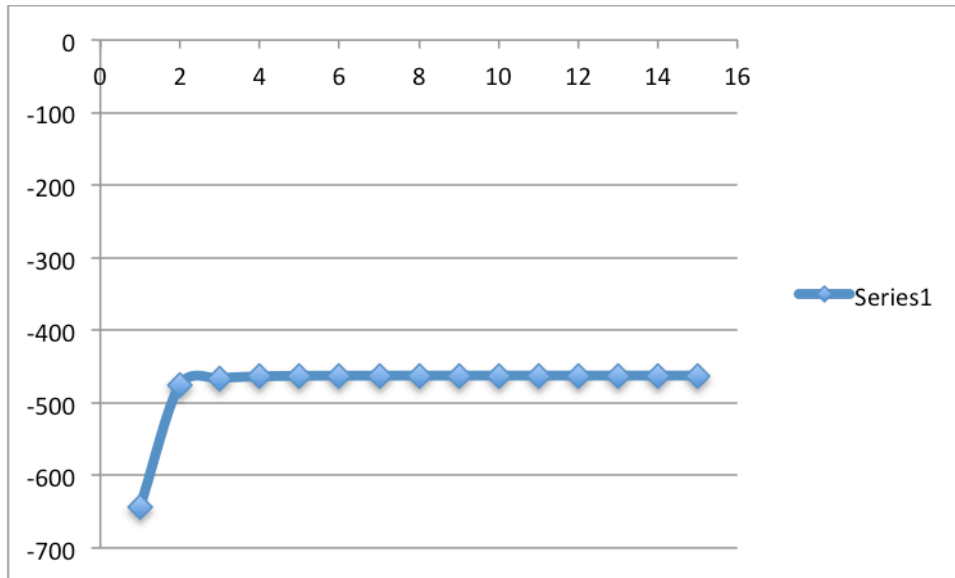
P(w=0|g=0) = 0.744626 P(w=1|g=0) = 0.255374

P(w=0|g=1) = 0.479256 P(w=1|g=1) = 0.520744

Height(h) given gender(g) table

P(h=0|g=0) = 0.687979 P(h=1|g=0) = 0.312021

P(h=0|g=1) = 0.338659 P(h=1|g=0) = 0.661341

The starting points of the learning:
P(gender=M) =0.42
P(w=0|g=0) = 0.68
P(w=0|g=1) = 0.19
P(h=0|g=0) = 0.23
P(h=0|g=1) = 0.63
The final conditional probability tables for hw2dataset_50.txt is:
Gender table
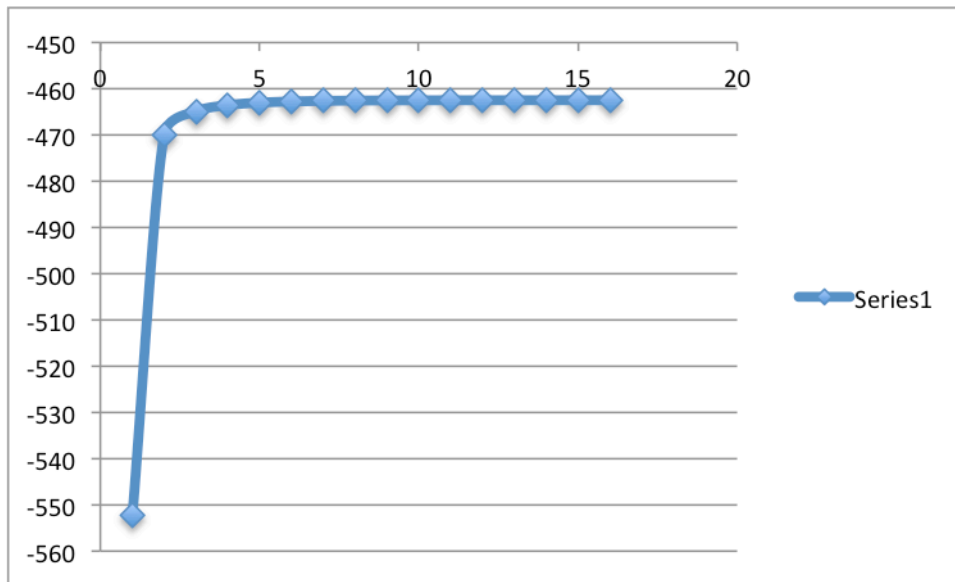P(gender=M) = 0.66234 P(gender=F) = 0.33766

Weight(w) given gender(g) table
P(w=0|g=0) = 0.744551 P(w=1|g=0) = 0.255449
P(w=0|g=1) = 0.479341 P(w=1|g=1) = 0.520659

Height(h) given gender(g) table
P(h=0|g=0) = 0.68802 P(h=1|g=0) = 0.31198
P(h=0|g=1) = 0.338497 P(h=1|g=0) = 0.661503

The final conditional probability tables for hw2dataset_50.txt is:
Gender table
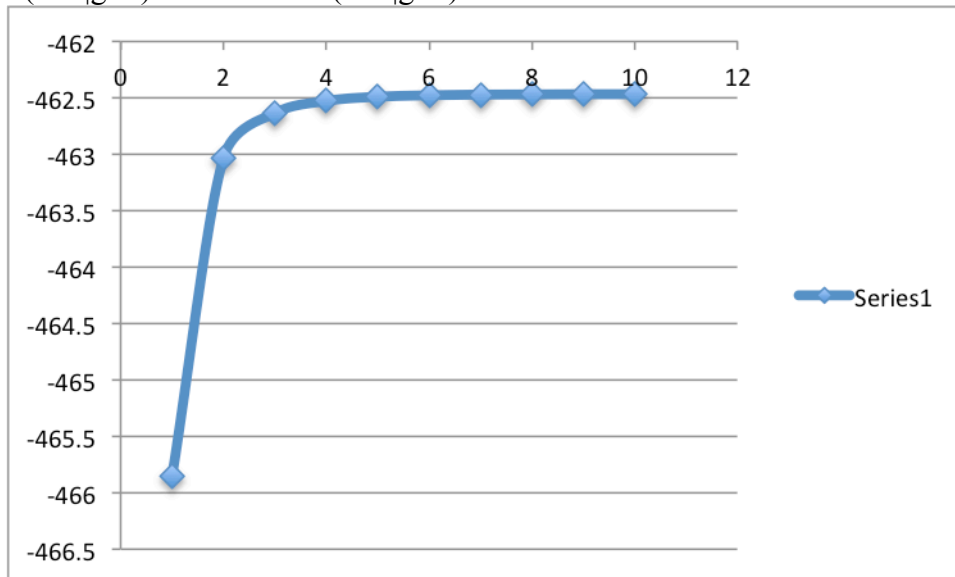P(gender=M) = 0.662702 P(gender=F) = 0.337298

Weight(w) given gender(g) table
P(w=0|g=0) = 0.744231 P(w=1|g=0) = 0.255769
P(w=0|g=1) = 0.479684 P(w=1|g=1) = 0.520316

Height(h) given gender(g) table
P(h=0|g=0) = 0.688183 P(h=1|g=0) = 0.311817
P(h=0|g=1) = 0.337801 P(h=1|g=0) = 0.662199

Data set with missing rate being 70%

The starting points of the learning:
P(gender=M) =0.67
P(w=0|g=0) = 0.62
P(w=0|g=1) = 0.9
P(h=0|g=0) = 0.13
P(h=0|g=1) = 0.42
The final conditional probability tables for hw2dataset_70.txt is:
Gender table
P(gender=M) = 0.690218 P(gender=F) = 0.309782

Weight(w) given gender(g) table
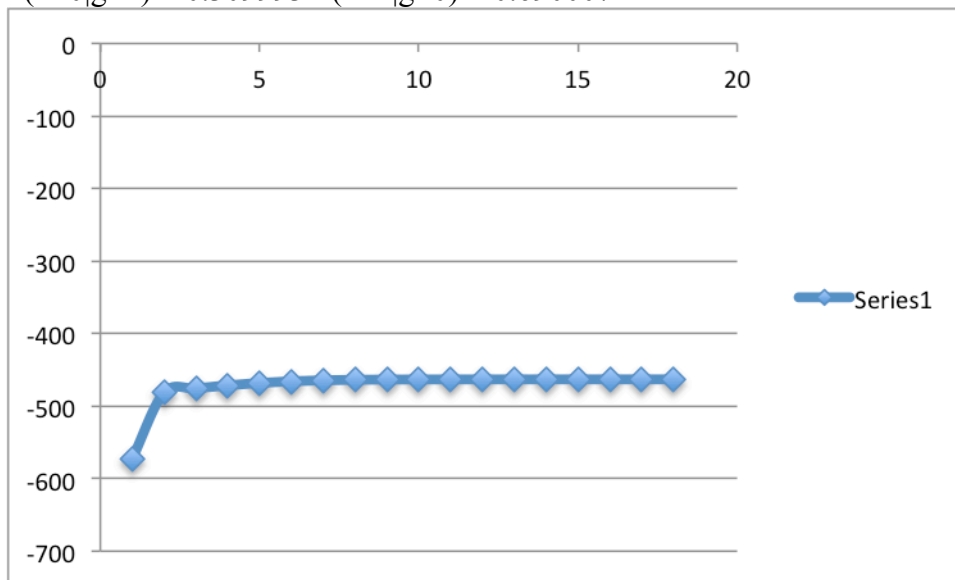P(w=0|g=0) = 0.829367 P(w=1|g=0) = 0.170633
P(w=0|g=1) = 0.4279 P(w=1|g=1) = 0.5721

Height(h) given gender(g) table
P(h=0|g=0) = 0.657719 P(h=1|g=0) = 0.342281
P(h=0|g=1) = 0.309993 P(h=1|g=0) = 0.690007

The starting points of the learning:
P(gender=M) =0.7
P(w=0|g=0) = 0.84
P(w=0|g=1) = 0.01
P(h=0|g=0) = 0.22
P(h=0|g=1) = 0.69
The final conditional probability tables for hw2dataset_70.txt is:
Gender table
P(gender=M) = 0.691768 P(gender=F) = 0.308232

Weight(w) given gender(g) table
P(w=0|g=0) = 0.829539 P(w=1|g=0) = 0.170461
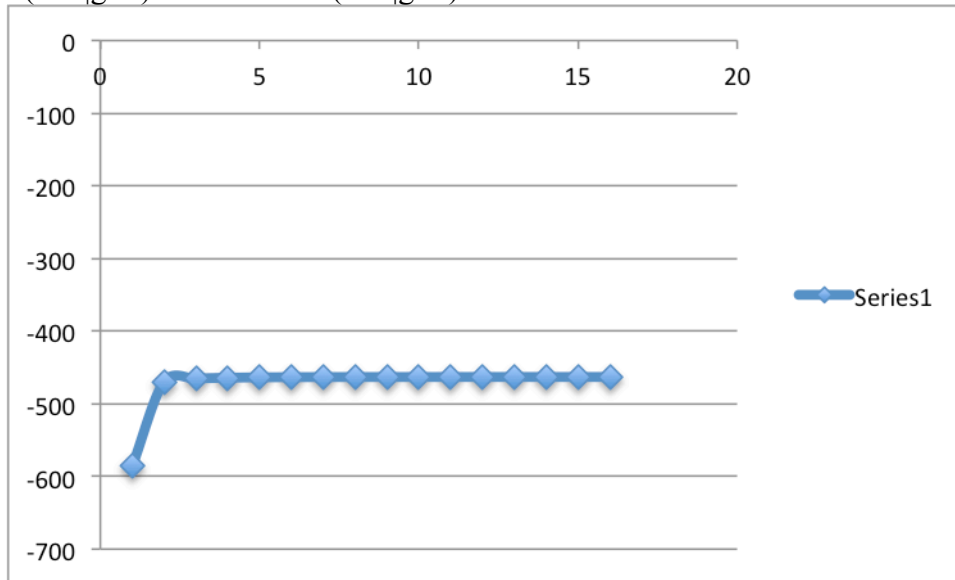
P(w=0|g=1) = 0.425494 P(w=1|g=1) = 0.574506

Height(h) given gender(g) table
P(h=0|g=0) = 0.656748 P(h=1|g=0) = 0.343252
P(h=0|g=1) = 0.310424 P(h=1|g=0) = 0.689576

The starting points of the learning:
P(gender=M) =0.63
P(w=0|g=0) = 0.24
P(w=0|g=1) = 0.03
P(h=0|g=0) = 0.02
P(h=0|g=1) = 0.95
The final conditional probability tables for hw2dataset_70.txt is:
Gender table
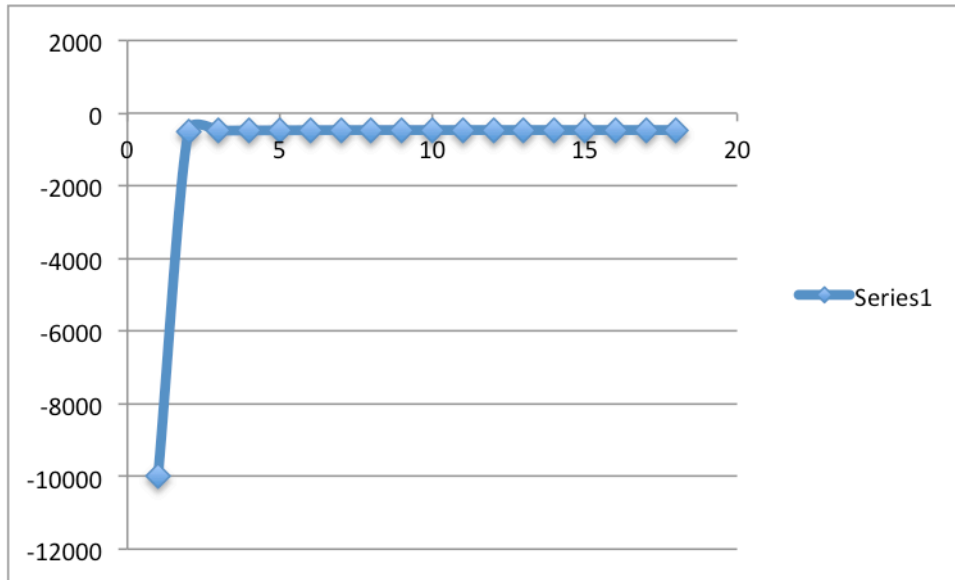P(gender=M) = 0.690471 P(gender=F) = 0.309529

Weight(w) given gender(g) table
P(w=0|g=0) = 0.829383 P(w=1|g=0) = 0.170617
P(w=0|g=1) = 0.427537 P(w=1|g=1) = 0.572463

Height(h) given gender(g) table
P(h=0|g=0) = 0.657549 P(h=1|g=0) = 0.342451
P(h=0|g=1) = 0.310089 P(h=1|g=0) = 0.689911

The starting points of the learning:
P(gender=M) =0.65
P(w=0|g=0) = 0.21
P(w=0|g=1) = 0.26
P(h=0|g=0) = 0.57
P(h=0|g=1) = 0.13
The final conditional probability tables for hw2dataset_70.txt is:
Gender table
P(gender=M) = 0.689193 P(gender=F) = 0.310807

Weight(w) given gender(g) table
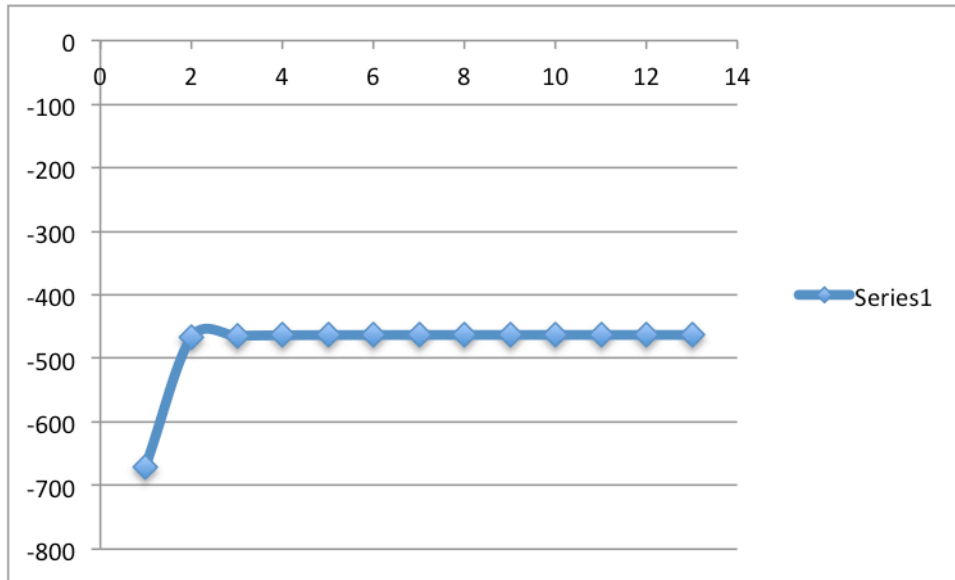P(w=0|g=0) = 0.829232 P(w=1|g=0) = 0.170768
P(w=0|g=1) = 0.429524 P(w=1|g=1) = 0.570476

Height(h) given gender(g) table
P(h=0|g=0) = 0.658431 P(h=1|g=0) = 0.341569
P(h=0|g=1) = 0.309562 P(h=1|g=0) = 0.690438

The starting points of the learning:
P(gender=M) =0.25
P(w=0|g=0) = 0.24
P(w=0|g=1) = 0.46
P(h=0|g=0) = 0.58
P(h=0|g=1) = 0.92
The final conditional probability tables for hw2dataset_70.txt is:
Gender table
P(gender=M) = 0.689672 P(gender=F) = 0.310328

Weight(w) given gender(g) table
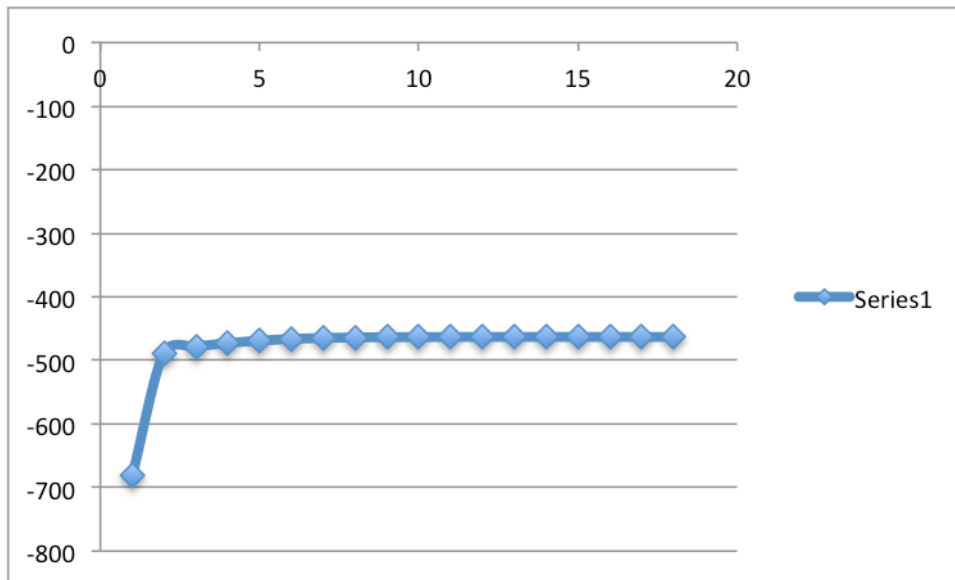P(w=0|g=0) = 0.829321 P(w=1|g=0) = 0.170679
P(w=0|g=1) = 0.42871 P(w=1|g=1) = 0.57129

Height(h) given gender(g) table
P(h=0|g=0) = 0.658064 P(h=1|g=0) = 0.341936
P(h=0|g=1) = 0.30984 P(h=1|g=0) = 0.69016

The final conditional probability tables for hw2dataset_70.txt is:
Gender table
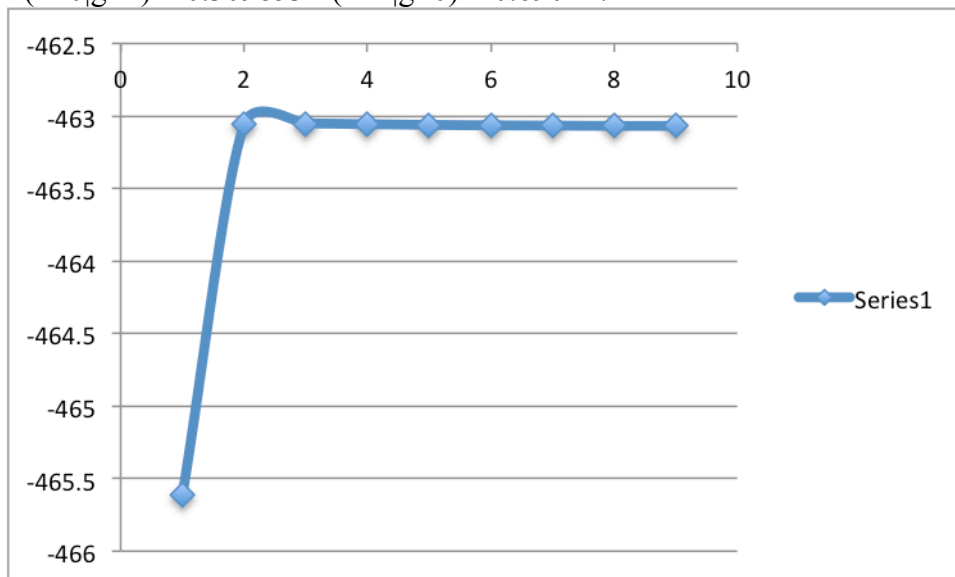P(gender=M) = 0.690582 P(gender=F) = 0.309418

Weight(w) given gender(g) table
P(w=0|g=0) = 0.829364 P(w=1|g=0) = 0.170636
P(w=0|g=1) = 0.427436 P(w=1|g=1) = 0.572564

Height(h) given gender(g) table
P(h=0|g=0) = 0.657599 P(h=1|g=0) = 0.342401
P(h=0|g=1) = 0.309853 P(h=1|g=0) = 0.690147

Data set with missing rate being 100%

The starting points of the learning:
P(gender=M) =0.29
P(w=0|g=0) = 0.18
P(w=0|g=1) = 0.92
P(h=0|g=0) = 0.21
P(h=0|g=1) = 0.75
The final conditional probability tables for hw2dataset_100.txt is:
Gender table
P(gender=M) = 0.298406 P(gender=F) = 0.701594

Weight(w) given gender(g) table
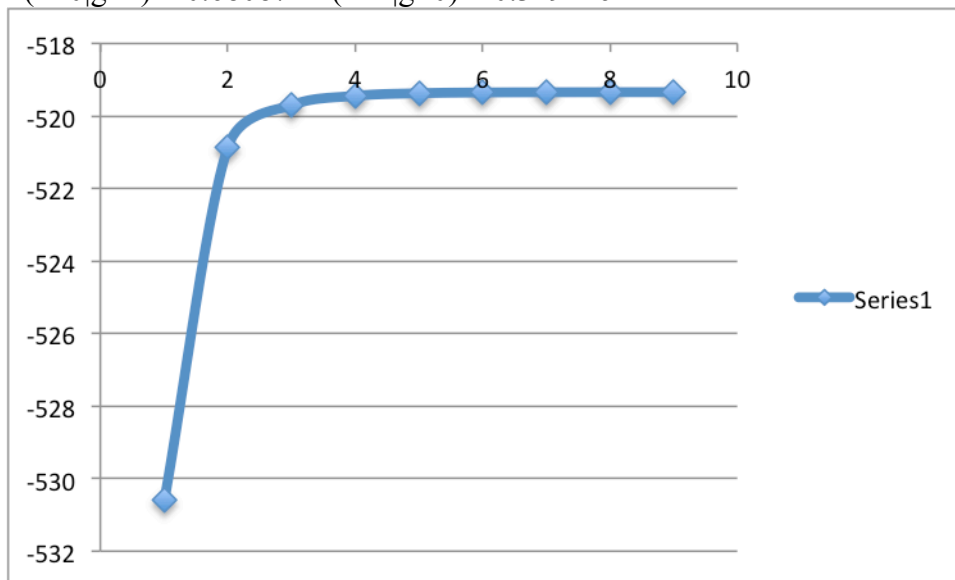P(w=0|g=0) = 0.205357 P(w=1|g=0) = 0.794643
P(w=0|g=1) = 0.860498 P(w=1|g=1) = 0.139502

Height(h) given gender(g) table
P(h=0|g=0) = 0.393098 P(h=1|g=0) = 0.606902
P(h=0|g=1) = 0.680874 P(h=1|g=0) = 0.319126

The starting points of the learning:
P(gender=M) =0.26
P(w=0|g=0) = 0.27
P(w=0|g=1) = 0.37
P(h=0|g=0) = 0.35
P(h=0|g=1) = 0.19
The final conditional probability tables for hw2dataset_100.txt is:
Gender table
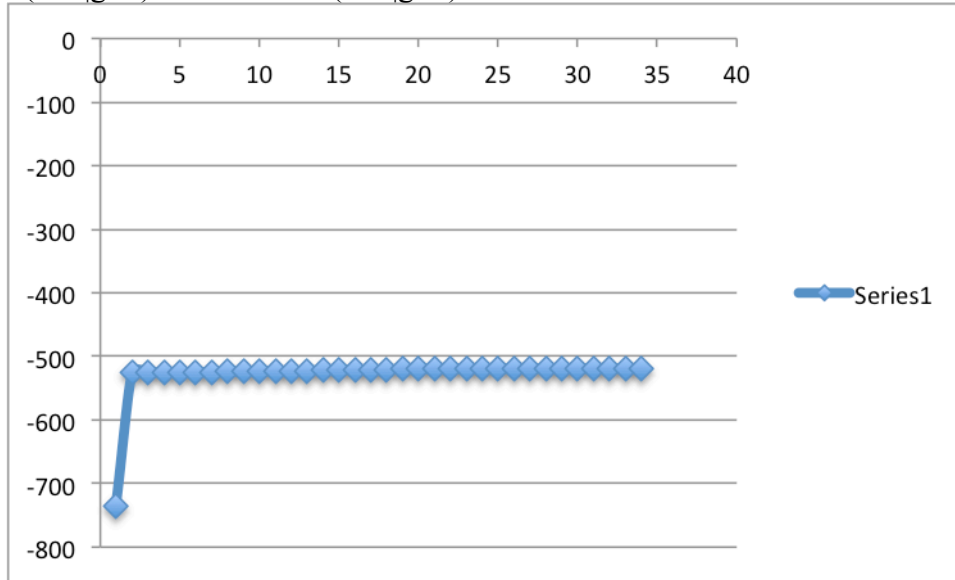P(gender=M) = 0.341385 P(gender=F) = 0.658615

Weight(w) given gender(g) table
P(w=0|g=0) = 0.922035 P(w=1|g=0) = 0.0779652

P(w=0|g=1) = 0.531769 P(w=1|g=1) = 0.468231

Height(h) given gender(g) table
P(h=0|g=0) = 0.8863 P(h=1|g=0) = 0.1137
P(h=0|g=1) = 0.444009 P(h=1|g=0) = 0.555991

The starting points of the learning:
P(gender=M) =0.24
P(w=0|g=0) = 0.08
P(w=0|g=1) = 0.17
P(h=0|g=0) = 0.56
P(h=0|g=1) = 0.72
The final conditional probability tables for hw2dataset_100.txt is:
Gender table
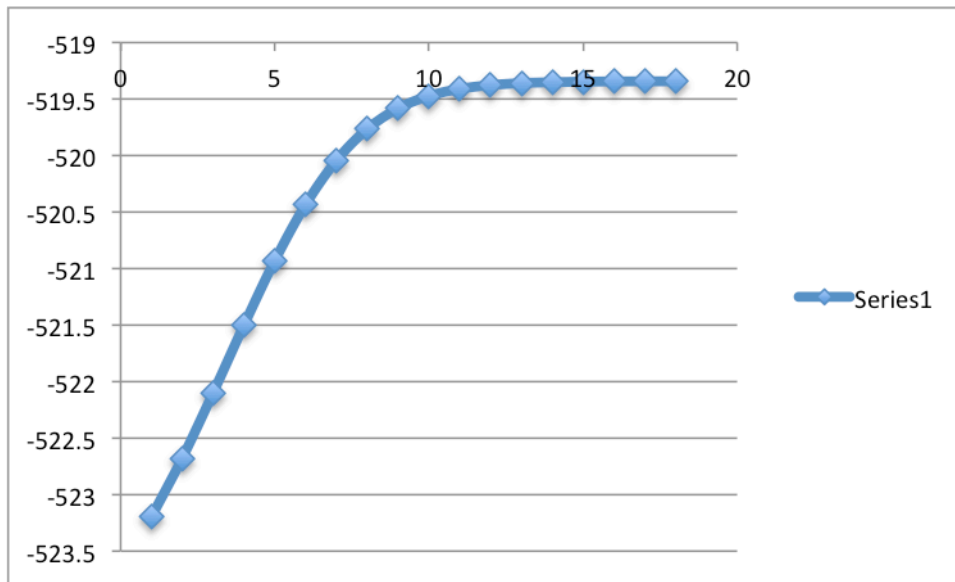P(gender=M) = 0.216463 P(gender=F) = 0.783537

Weight(w) given gender(g) table
P(w=0|g=0) = 0.283082 P(w=1|g=0) = 0.716918
P(w=0|g=1) = 0.77051 P(w=1|g=1) = 0.22949

Height(h) given gender(g) table
P(h=0|g=0) = 0.224495 P(h=1|g=0) = 0.775505
P(h=0|g=1) = 0.697357 P(h=1|g=0) = 0.302643

The starting points of the learning:
P(gender=M) =0.17
P(w=0|g=0) = 0.48
P(w=0|g=1) = 0.19
P(h=0|g=0) = 0.83
P(h=0|g=1) = 0.51
The final conditional probability tables for hw2dataset_100.txt is:
Gender table
P(gender=M) = 0.309605 P(gender=F) = 0.690395

Weight(w) given gender(g) table
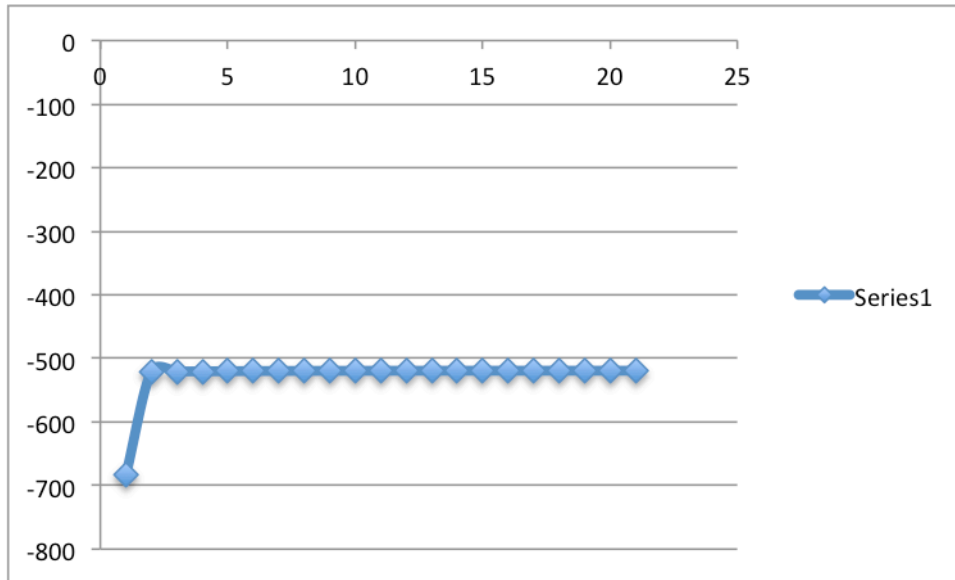P(w=0|g=0) = 0.937714 P(w=1|g=0) = 0.062286
P(w=0|g=1) = 0.542703 P(w=1|g=1) = 0.457297

Height(h) given gender(g) table
P(h=0|g=0) = 0.911686 P(h=1|g=0) = 0.0883142
P(h=0|g=1) = 0.452984 P(h=1|g=0) = 0.547016

The starting points of the learning:
P(gender=M) =0.47
P(w=0|g=0) = 0.94
P(w=0|g=1) = 0.93
P(h=0|g=0) = 0.76
P(h=0|g=1) = 0.95
The final conditional probability tables for hw2dataset_100.txt is:
Gender table
P(gender=M) = 0.550264 P(gender=F) = 0.449736

Weight(w) given gender(g) table
P(w=0|g=0) = 0.518737 P(w=1|g=0) = 0.481263
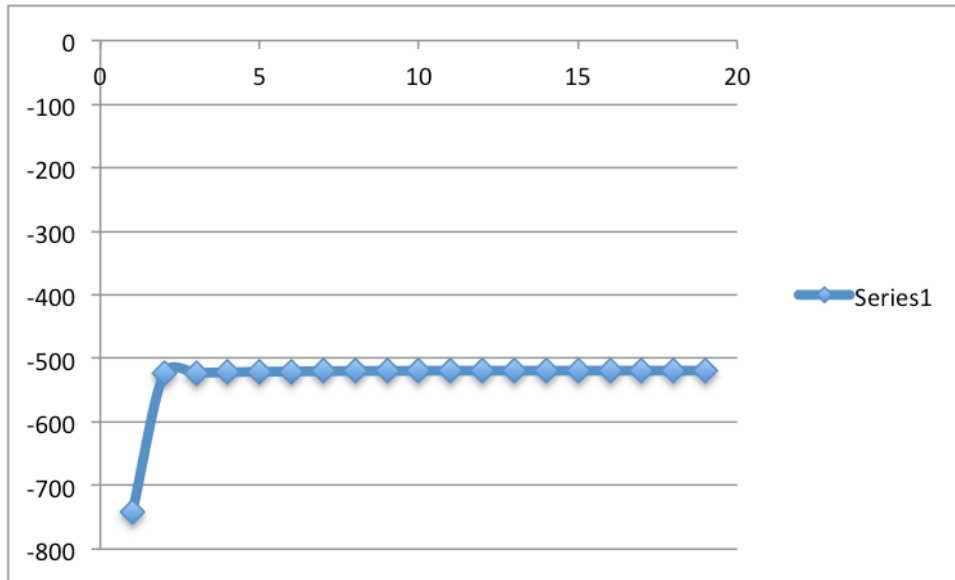P(w=0|g=1) = 0.843957 P(w=1|g=1) = 0.156043

Height(h) given gender(g) table
P(h=0|g=0) = 0.377 P(h=1|g=0) = 0.623
P(h=0|g=1) = 0.861728 P(h=1|g=0) = 0.138272

The final conditional probability tables for hw2dataset_100.txt is:
Gender table
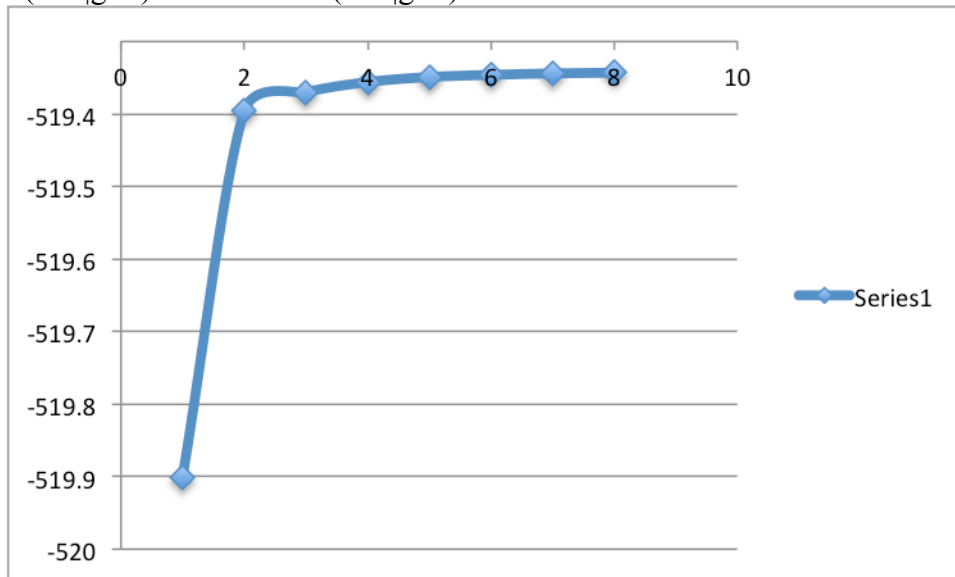P(gender=M) = 0.696858 P(gender=F) = 0.303142

Weight(w) given gender(g) table
P(w=0|g=0) = 0.796089 P(w=1|g=0) = 0.203911
P(w=0|g=1) = 0.363655 P(w=1|g=1) = 0.636345

Height(h) given gender(g) table
P(h=0|g=0) = 0.724454 P(h=1|g=0) = 0.275546
P(h=0|g=1) = 0.297413 P(h=1|g=0) = 0.702587

Analysis

1. Do multiple starting points help in finding better solutions?

Yes, multiple starting points help in finding better solutions. Because for EM Algorithm, it only guarantees local maxima, there might exit a better global maxima. From multiple starting points result above, though in general the solution falls into the same range, but the bad starting points occur, the solution is not accurate.

2. Do some of the different solutions have the same likelihood scores?

Yes. For example, for data set with missing rates being 30%, the third and fourth solutions are different but they have the same likelihood scores.

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

Generally, when data missing rate is low, it takes less time to converge and the results are more accurate. The higher the missing rate is, the less accurate results are. Worst case is when missing rate is 100%, there are many solutions.

4. When starting points are in reasonably guessing, they are faster to converge.

5. Avoiding bad local maxima by multiple restarts and early pruning of unpromising starting points.