Start Point: w = (0)

Log-likelihood: -970.4061

Difference < 0.001:

Log-likelihood: -151.5325

Number of iterations: 5989

Difference < 0.0001:

Log-likelihood: -147.9591

Number of iterations: 16293

In the 17000th iteration,

w =

-1.3637 -1.4414 -2.0814 -0.9597 -1.7514 -0.4850 0.8857 1.9567

0.8428 0.0206 0.9162 -0.6138 -0.1233 0.5922 -1.5064 0.0034

2.8400 1.1854 1.0412 0.3395 0.2824 -2.1745 -3.1405 -3.6695

2.2742 0.7275 1.6632 -0.5657 -1.5898 -0.5549 0.4115 -0.2380

0.2042 0.2624 0.2318 -0.9280 -0.1605 -0.1029 -0.7062 -0.1677

1.2957 -0.9795 0.5357 0.6847 0.4476 -0.7437 0.0308 -1.8091

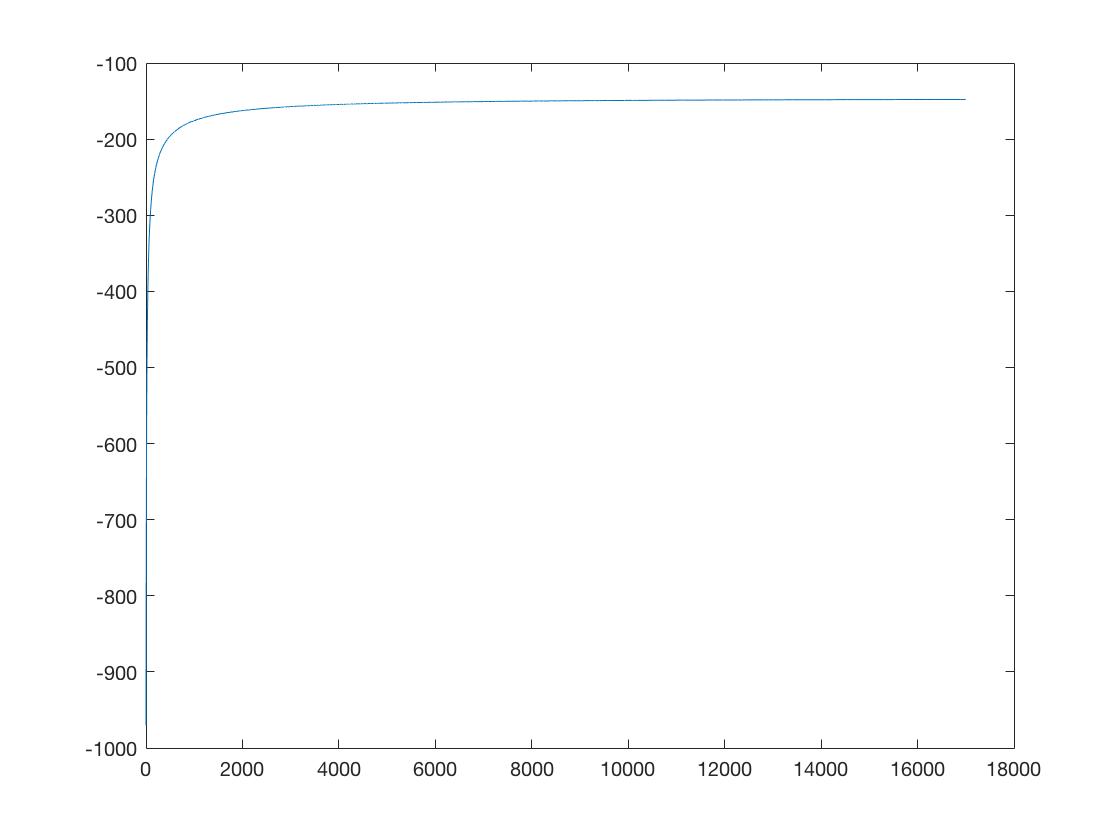
0.5052 -0.2876 1.0806 0.8777 -0.0174 -0.2429 0.5469 -1.6282

0.3255 0.3474 -0.3624 4.5096 0.4771 0.4901 0.0107 -0.6354

5.5(a)

Algorithm used: Gradient Ascent

Here is the plot of log-likelihood on number of iterations:



Note: The difference between consecutive values of log-likelihood is less than 0.0001 after the 16293th iteration.

Percent error rate on these two training files:

52/1400 ≈ 3.71%

My weight vector w is:

w =

-1.3637 -1.4414 -2.0814 -0.9597 -1.7514 -0.4850 0.8857 1.9567

0.8428 0.0206 0.9162 -0.6138 -0.1233 0.5922 -1.5064 0.0034

2.8400 1.1854 1.0412 0.3395 0.2824 -2.1745 -3.1405 -3.6695

2.2742 0.7275 1.6632 -0.5657 -1.5898 -0.5549 0.4115 -0.2380

0.2042 0.2624 0.2318 -0.9280 -0.1605 -0.1029 -0.7062 -0.1677

1.2957 -0.9795 0.5357 0.6847 0.4476 -0.7437 0.0308 -1.8091

0.5052 -0.2876 1.0806 0.8777 -0.0174 -0.2429 0.5469 -1.6282

0.3255 0.3474 -0.3624 4.5096 0.4771 0.4901 0.0107 -0.6354

(b)

Percent error rate on the two testing files:

42/800 ≈ 5.25%