6. (a) The likelihood function converges to a certain value after about 5000 iterations with a learning rate of 0.001.

The log likelihood function value in the first 9 iterations are listed below.

```
log:  -970.406052784  log:  -701.315523713  log:  -1637.42075794
log:  -5535.01075562  log:  -1555.18897641  log:  -3766.53527502
log:  -1727.2463233   log:  -2537.32923792  log:  -1174.17562683
log:  -1348.51929809  log:  -755.853656746  log:  -715.744051387
```

Actually, during the first several iterations, the value is bouncing back and forth. Then start from the 12th iteration, the function value starts to gradient descent in one direction.

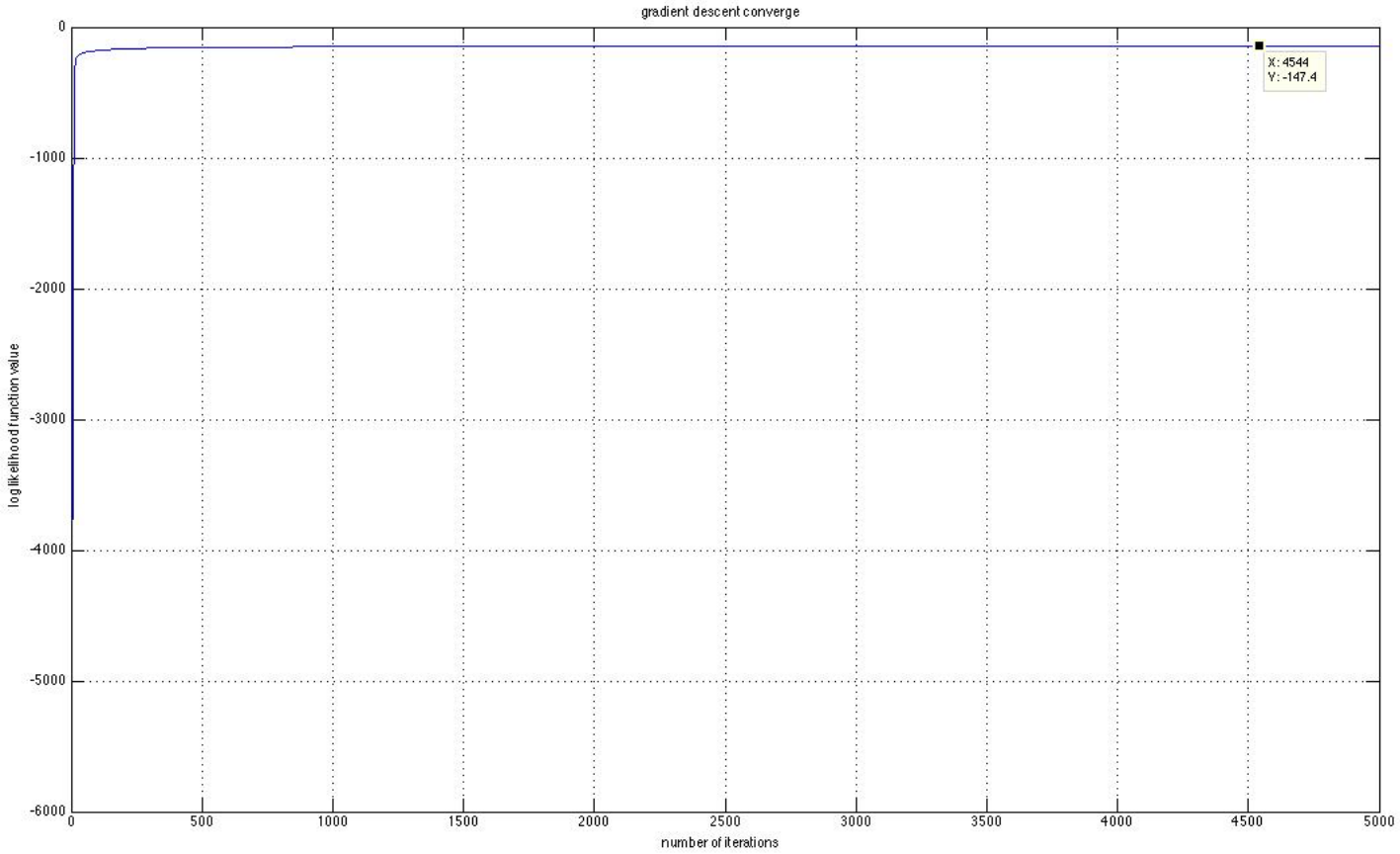
In the last 2000 iterations, the value converges very slowly towards -147.3495. Here I just list several values from the last iterations.

```
log:  -147.349714556
log:  -147.349700592
log:  -147.349686643
log:  -147.349672707
log:  -147.349658786
log:  -147.349644878
log:  -147.349630985
log:  -147.349617105
log:  -147.349603239
log:  -147.349589388
log:  -147.34957555
log:  -147.349561726
log:  -147.349547916
log:  -147.34953412
```

The weight vector is:

```
w:  [-1.53301403 -1.48429062 -2.22079829 -0.97105807 -1.86732557 -0.4883959
  0.88138338  2.04514895  1.03108567 -0.04839772  1.01757989 -0.66309329
 -0.14786594  0.63340024 -1.57905459  0.02907137  2.95291747  1.27848733
  1.01640661  0.34069904  0.26199445 -2.27179045 -3.27406204 -3.85295161
  2.42791959  0.75005343  1.79097086 -0.58383879 -1.74188449 -0.55677888
  0.46715707 -0.29512893  0.18246235  0.23393425  0.27509105 -1.0230082
 -0.09269407 -0.07330668 -0.78303051 -0.14496728  1.35813881 -0.98776292
  0.50471822  0.7916843   0.41091148 -0.80297218  0.01772689 -1.91028338
  0.60648545 -0.35100621  1.17989136  0.91618812 -0.12399735 -0.19726054
  0.61538495 -1.75124618  0.39788889  0.37184725 -0.87843639  5.41213072
  0.36607671  0.43560885  0.0293102  -0.66265953]
```

The convergence plot is given.



(b)

```
error on training data 3:  0.037142857
error on training data 5:  0.034285714
overall error on training data: 0.035714285

error on test data 3:  0.055
error on test data 5:  0.05
overall error on test data: 0.0525
```