5.5

$$L = logP(\{y_t, t = 1,2, \dots T\}|\{x_t, t = 1,2, \dots T\})$$

$$L = logP \prod_{t=1}^{T} P(y_t|x_t)$$

…………………………………………………………………(IID) identical independent data

$$L = \sum_{t=1}^{T} logP(y_t|x_t)$$

$$L = \sum_{t=1}^{T} \log\left(P(y_t = 1|x_t)^{y_t} P(y_t = 0|x_t)^{(1-y_t)}\right)$$

………………………………………………………………………………$y_t = 0,1$

Where

$$P(y_t = 1|x_t) = \sigma(\vec{w} * \vec{x_t})$$

$$P(y_t = 0|x_t) = 1 - \sigma(\vec{w} * \vec{x_t}) = \sigma(-\vec{w} * \vec{x_t})$$

Take it into this, we can get

$$L = \sum_{t=1}^{T} (y_t \log\sigma(\vec{w} * \vec{x_t}) + (1 - y_t)\log\sigma(-\vec{w} * \vec{x_t}))$$

So we have

$$\frac{\partial L}{\partial w_a} = \sum_{t=1}^{T} (y_t \frac{\sigma'(\vec{w} * \vec{x_t})}{\sigma(\vec{w} * \vec{x_t})} x_{ta} + (1 - y_t) \frac{\sigma'(-\vec{w} * \vec{x_t})}{\sigma(-\vec{w} * \vec{x_t})} (-1) x_{ta})$$

Where

$$\sigma'(\vec{w} * \vec{x_t}) = \sigma(\vec{w} * \vec{x_t})\sigma(-\vec{w} * \vec{x_t})$$

$$\frac{\partial L}{\partial w_a} = \sum_{t=1}^{T} (y_t \frac{\sigma(\vec{w} * \vec{x_t})\sigma(-\vec{w} * \vec{x_t})}{\sigma(\vec{w} * \vec{x_t})} x_{ta} + (1 - y_t) \frac{\sigma(\vec{w} * \vec{x_t})\sigma(-\vec{w} * \vec{x_t})}{\sigma(-\vec{w} * \vec{x_t})} (-1) x_{ta})$$

$$\frac{\partial L}{\partial w_a} = \sum_{t=1}^{T} (y_t \sigma(-\vec{w} * \vec{x_t}) x_{ta} + (1 - y_t)\sigma(\vec{w} * \vec{x_t})(-1) x_{ta})$$

$$\frac{\partial L}{\partial w_a} = \sum_{t=1}^{T} (y_t (1 - \sigma(\vec{w} * \vec{x_t})) - (1 - y_t)\sigma(\vec{w} * \vec{x_t})) x_{ta}$$

$$\frac{\partial L}{\partial w_a} = \sum_{t=1}^{T} (y_t - y_t \sigma(\vec{w} * \vec{x_t}) - \sigma(\vec{w} * \vec{x_t}) + y_t \sigma(\vec{w} * \vec{x_t})) x_{ta}$$

$$\frac{\partial L}{\partial w_a} = \sum_{t=1}^{T} (y_t - \sigma(\vec{w} * \vec{x_t})) x_{ta}$$

Using Gradient ascent, then we have:

$$\vec{w} \; < - \, \vec{w} + \eta \left(\frac{\partial L}{\partial w_a} \right)$$

As recommended, we set

$$\eta = \frac{0.2}{T} = \frac{1}{5T}$$

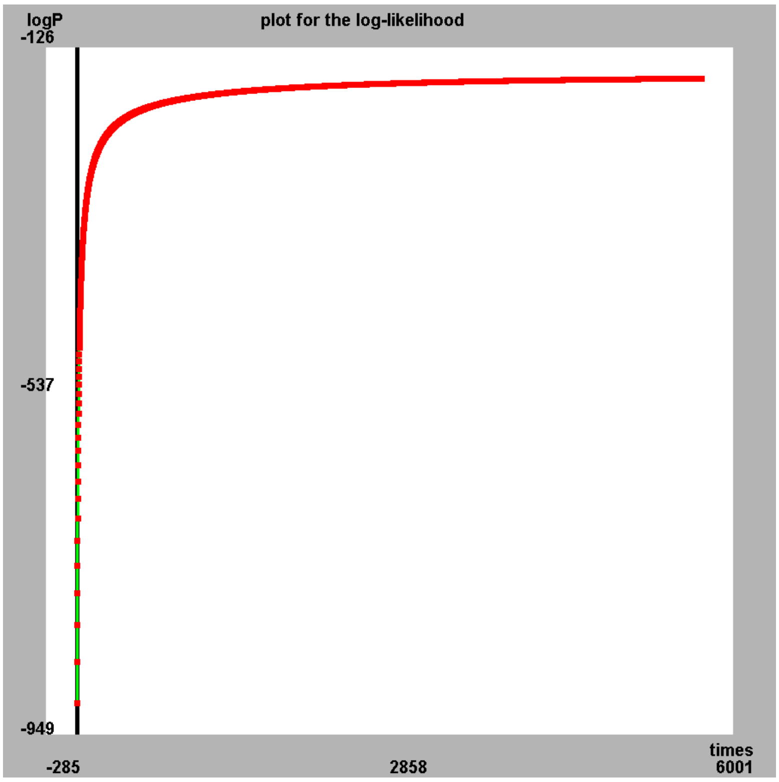
So we have

$$\vec{w} \; < - \, \vec{w} + \frac{1}{5T} \left(\frac{\partial L}{\partial w_a} \right)$$

And we start at

$$\vec{w}_0 = (0,0,0 \ldots .0)$$

After running the program, we can get, it converged



Log-likelihood on Several iterations

CSE250A0505

```
5590___-163.7728007141486____7.368468007697004
============================================
5600___-163.7622991163491____7.3383163711529065
============================================
5610___-163.75184046666126____7.308317568965711
============================================
5620___-163.74142454754818____7.2784706310697525
============================================
5630___-163.7310511428528____7.248774595018466
============================================
5640___-163.72072003778558____7.21922850591162
============================================
5650___-163.7104310189158____7.189831416323277
============================================
5660___-163.70018387416047____7.160582386232594
============================================
5670___-163.68997839277262____7.1314804829528295
============================================
5680___-163.679814365332____7.10252478106342
============================================
5690___-163.66969158373504____7.073714362339671
============================================
5700___-163.65960984118297____7.045048315687634
============================================
5710___-163.64956893217382____7.016525737075729
```

Percent error:

```
====================error rate in the training data=
percent error rate___0.03642857142857143
error count___51
total count(T)___1400
====================error rate in the testing data==
```

Solution for the weight vector in an 8*8 matrix:

```
5710___-163.64956893217382____7.016525737075729
===================weight matrix 8*8=======================
-0.888943	-1.430894	-1.157044	-1.138776	-0.731800	-0.820299	0.810885	1.729220
0.041149	-0.110997	0.216132	-0.069145	-0.374167	0.720662	-1.255910	-1.297928
3.363476	1.364344	1.384051	0.203669	0.668948	-1.974978	-2.392744	-2.485500
0.781593	0.394691	0.584120	-0.282440	-0.481714	-2.235679	0.378615	-0.028924
0.491810	1.076402	0.049990	-0.335482	-0.632241	-0.182845	-0.430495	-0.299055
1.145425	-0.199077	-0.323594	-0.084782	0.088347	-0.851867	0.796668	-1.467420
1.407552	-0.622698	1.265531	0.572139	0.419400	-0.313982	0.205782	-1.190904
0.548130	0.279823	0.882708	1.770618	0.480285	0.637082	0.568125	-0.474228
```

(b)

```
total count(T)___1400
====================error rate in the testing data
percent error rate___0.06125
error count___49
total count(T)___800
```

(c)

Source code:

In the source code, I use the plot package for the plotting, which is developed by myself from scratch, I attach it in the end