1.

(1) ∵
$$W^{new} = W^{old} - H^{-1} \nabla E(w)$$

$$\nabla_{w_j} E(w_1, w_2 \cdots w_k) = \sum_{n=1}^{N} (y_{nj} - t_{nj}) \phi_n \quad \begin{pmatrix} n \to \text{data number} \\ j \to \text{class number} \\ y \to \text{predict posterior} \\ t \to \text{real posterior} \end{pmatrix}$$

$$H_{kj}$$
$$\Downarrow$$
$$\nabla_{w_k} \nabla_{w_j} E(w_1, w_2 \cdots w_k) = \sum_{n=1}^{N} y_{nk} (I_{kj} - y_{nj}) \phi_n \phi_n^T$$
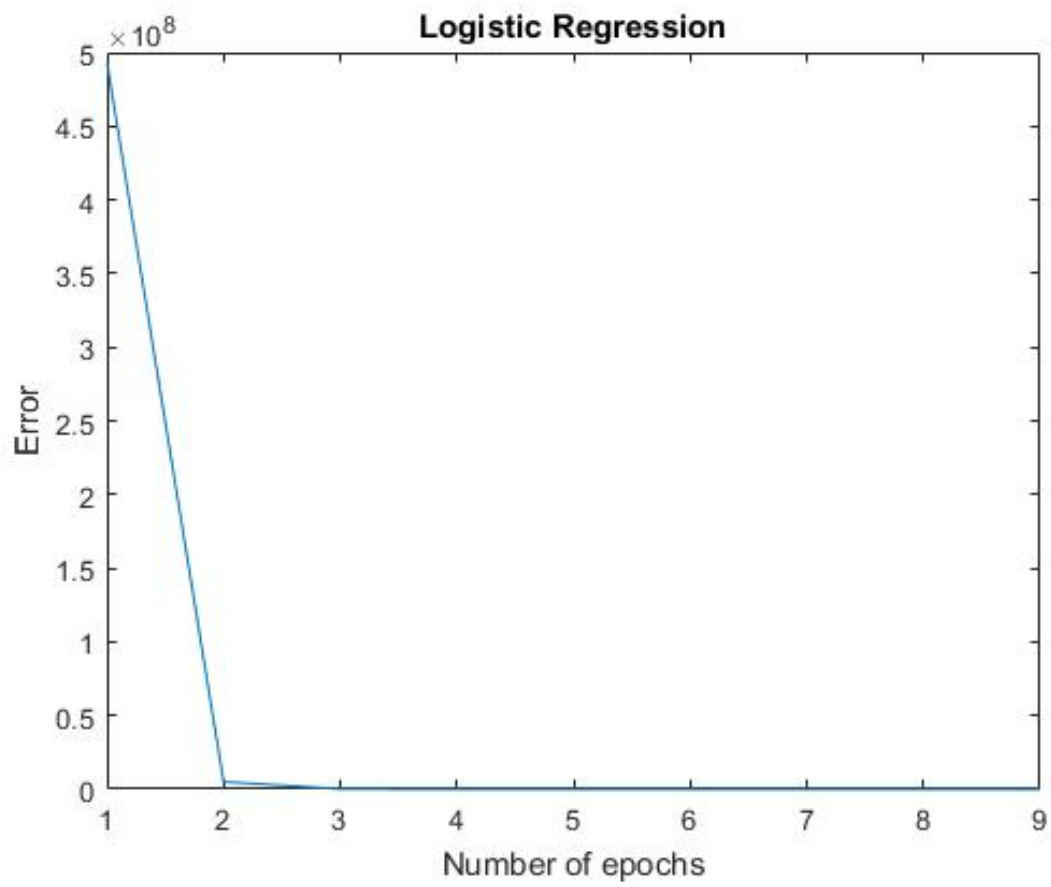
$$\Rightarrow W_1^{new} = W_1^{old} - H_{11}^{-1} \nabla_{w_1} E(w)$$

$$W_2^{new} = W_2^{old} - H_{22}^{-1} \nabla_{w_2} E(w)$$

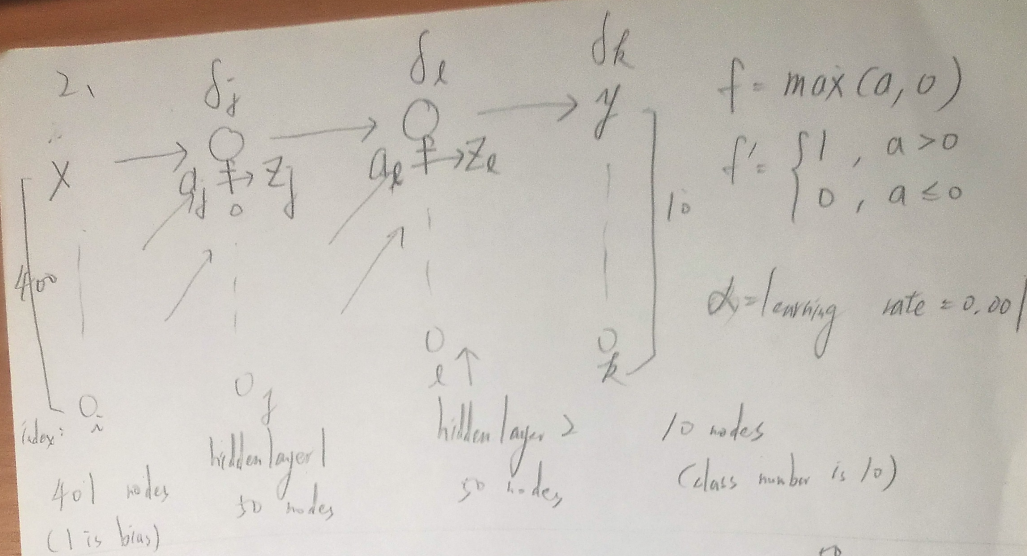$$W_3^{new} = W_3^{old} - H_{33}^{-1} \nabla_{w_3} E(w) \quad \#$$

(2)

$$P(C_1 | \phi) = y_1(\phi) = \frac{e^{a_1}}{e^{a_1} + e^{a_2} + e^{a_3}}$$

$\#$

Posterior

3.096e-19
6.4502e-17
6.2441e-16
7.878e-16
8.1919e-09
1.3642e-17
1.0507e-19
2.0657e-14
4.2345e-19
1.2338e-12
3.9453e-11
0.0018096
0.12885
2.0614e-06
7.5128e-21
1.3616e-18
2.9282e-14
4.4409e-20
2.8026e-13
0.60181
0.9559
0.99971
0.99669
0.99952
0.99991
1
0.99999
0.99999
1
1

2.

$$f = \max(0, 0)$$

$$f' = \begin{cases} 1, & a > 0 \\ 0, & a \leq 0 \end{cases}$$

$$\alpha = \text{learning rate} = 0.001$$

index:

401 nodes
(1 is bias)

hidden layer 1
50 nodes

hidden layer 2
50 nodes

10 nodes
(class number is 10)

$$a_j = \sum_i^{401} W_{ji} X_i \qquad a_\ell = \sum_{j=0}^{50} W_{\ell j} Z_j \qquad a_k = \sum_{\ell=0}^{50} W_{k\ell} Z_\ell$$

$$\Delta w_{ji} = -\alpha \frac{\partial E}{\partial w_{ji}} \qquad \Delta w_{\ell j} = -\alpha \frac{\partial E}{\partial w_{\ell j}} \qquad \Delta W_{k\ell} = -\alpha \frac{\partial E}{\partial w_{k\ell}}$$

$$= -\alpha \delta_j X_i \qquad\qquad = -\alpha \delta_\ell Z_j \qquad\qquad = -\alpha \delta_k Z_\ell$$

$$\delta_j = f'(a_j) \sum_\ell W_{\ell j} \delta_\ell \qquad \delta_\ell = f'(a_\ell) \sum_k W_{k\ell} \delta_k \qquad \delta_k = y_k - t_k$$

Train Error

Train Misclassification Rate

Test Error

Test Misclassification Rate