HW4 Tao Ni

2019年4月14日 星期日 下午9:26

Q1: a. $E(w_1, w_2, v|X) = -\sum_t r^t \log y^t + (1-r^t)\log(1-y^t)$, $y^t = \text{sigmoid}\left(\sum_{h=1}^{2} v_h z_h^t + v_0\right)$

$$\frac{\partial E}{\partial v_h} = -\sum_t \frac{\partial E}{\partial y^t}\cdot\frac{\partial y^t}{\partial v_h} = -\sum_t\left(\frac{r^t}{y^t} - \frac{1-r^t}{1-y^t}\right)\cdot\frac{\partial y^t}{\partial a}\cdot\frac{\partial a}{\partial v_h}$$

$$= -\sum_t\left(\frac{r^t}{y^t} - \frac{1-r^t}{1-y^t}\right)\cdot y^t(1-y^t)\cdot z_h^t$$

$$= -\sum_t (r^t - y^t) z_h^t$$

$$\therefore \Delta v_h = \eta\sum_t (r^t - y^t) z_h^t$$

$$\frac{\partial E}{\partial w_{1j}} = -\sum_t \frac{\partial E}{\partial y^t}\cdot\frac{\partial y^t}{\partial z_1^t}\cdot\frac{\partial z_1^t}{\partial w_{1j}} = -\sum_t (r^t - y^t)\cdot v_1\cdot z_1^t(1-z_1^t)\cdot x_j^t$$

$$\therefore \Delta w_{1j} = \eta\sum_t (r^t - y^t) v_1\cdot z_1^t(1-z_1^t) x_j^t$$

$$\frac{\partial E}{\partial w_{2j}} = -\sum_t \frac{\partial E}{\partial y^t}\cdot\frac{\partial y^t}{\partial z_2^t}\cdot\frac{\partial z_2^t}{\partial w_{2j}} = \begin{cases} -\sum_t (r^t - y^t)\cdot v_2\cdot 0.01 x_j^t \\ \quad \text{if } \sum_{j=0}^{2} w_{2j} x_j^t < 0,\ x_0^t = 1 \\ -\sum_t (r^t - y^t) v_2 x_j^t \quad \text{otherwise} \end{cases}$$

$$\therefore \Delta w_{2j} = \begin{cases} 0.01\eta\sum_t (r^t - y^t) v_2 x_j^t & \text{if } \sum_{j=0}^{2} w_{2j} x_j^t < 0,\ x_0^t = 1 \\ \eta\sum_t (r^t - y^t) v_2 x_j^t & \text{otherwise} \end{cases}$$

Updating Equations:

$$\begin{cases} v := v + \eta\Delta v, & \Delta v = \begin{pmatrix} \Delta v_0 \\ \Delta v_1 \\ \Delta v_2 \end{pmatrix} \\ w_1 := w_1 + \eta\Delta w_1, & \Delta w_1 = \begin{pmatrix} \Delta w_{10} \\ \Delta w_{11} \\ \Delta w_{12} \end{pmatrix} \\ w_2 := w_2 + \eta\Delta w_2, & \Delta w_2 = \begin{pmatrix} \Delta w_{20} \\ \Delta w_{21} \end{pmatrix} \end{cases}$$

$$\Delta w_{22}$$

b. $w = w_1 = w_2$

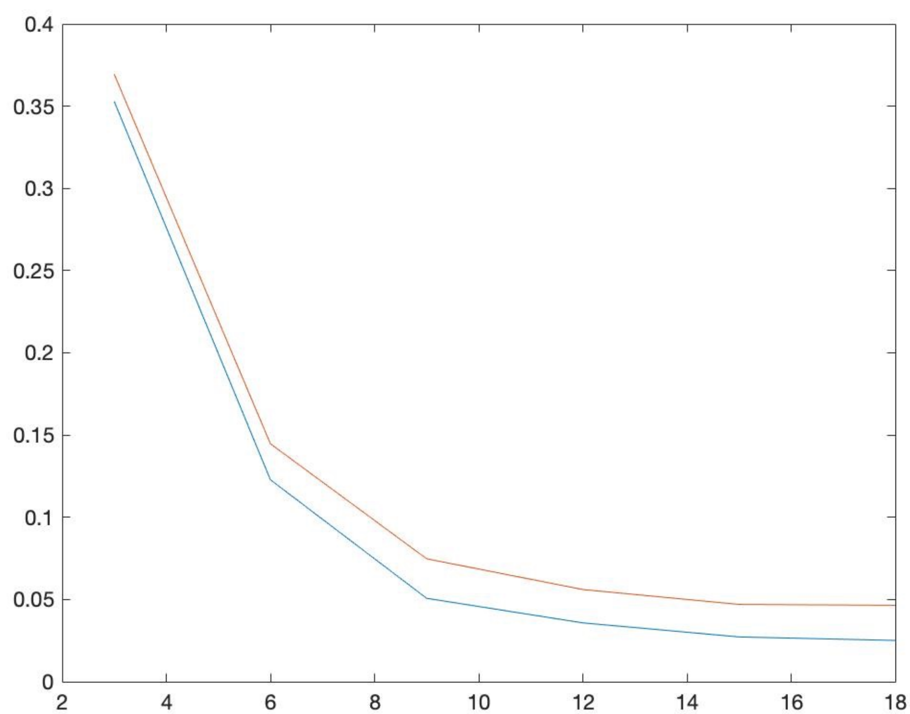
$$\frac{\partial E}{\partial v_h} = -\sum_t (r^t - y^t) z_h^t \text{ , same as above.}$$

$$\frac{\partial E}{\partial w_j} = -\sum_t \frac{\partial E}{\partial y^t} \left(\frac{\partial y^t}{\partial z_1^t} \cdot \frac{\partial z_1^t}{\partial w_j} + \frac{\partial y^t}{\partial z_2^t} \cdot \frac{\partial z_2^t}{\partial w_j} \right)$$

$$= \begin{cases} -\sum_t (r^t - y^t) \cdot \left(v_1 z_1^t (1 - z_1^t) x_j^t + 0.01 v_2 x_j^t \right) & \text{if } \sum_{j=0}^{2} w_j x_j^t < 0 \\ & x_0^t = 1 \\ -\sum_t (r^t - y^t) \cdot \left(v_1 z_1^t (1 - z_1^t) x_j^t + v_2 x_j^t \right) & \text{otherwise.} \end{cases}$$

$$\Delta w_j = -\eta \frac{\partial E}{\partial w_j}$$

2.a: Red - validation error; Blue - training error

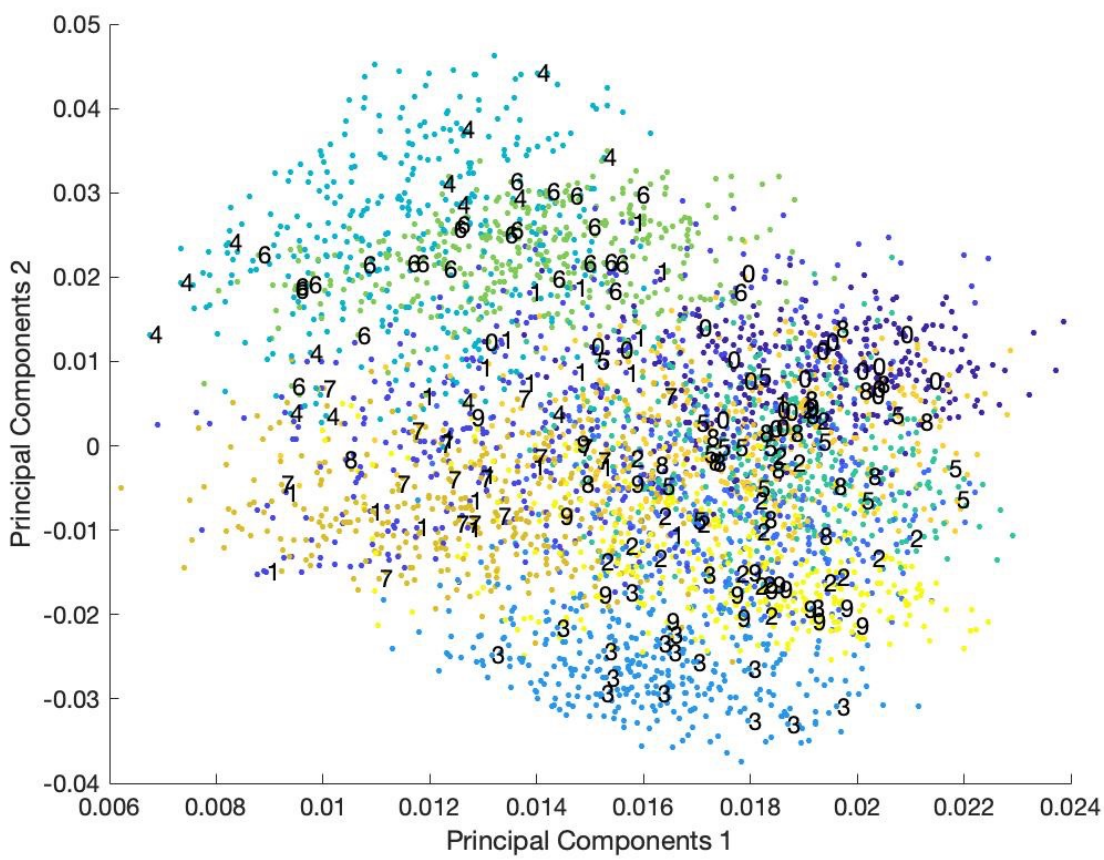


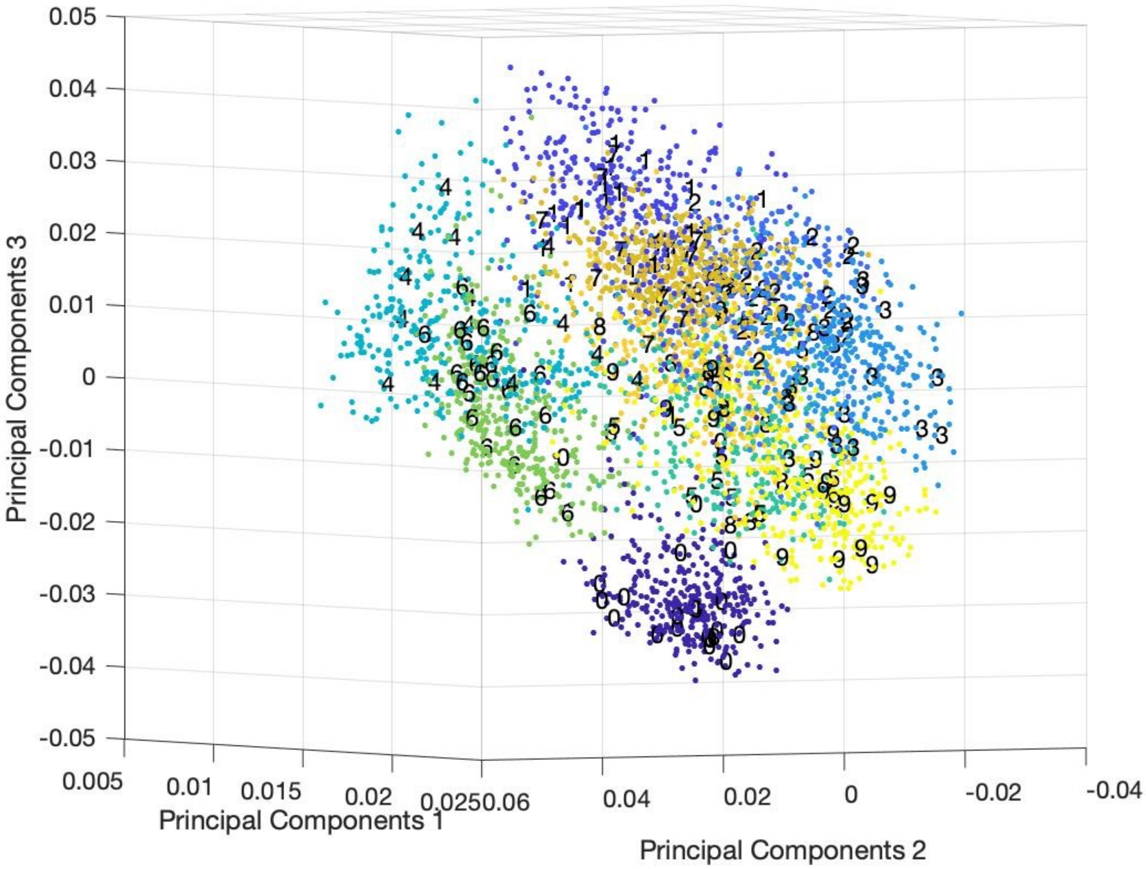
I would use 18 hidden units since it achieves both the minimum training error and validation error.

# of hidden units	3	6	9	12	15	18
Train	0.3529	0.1228	0.0507	0.0358	0.0272	0.0251
Validation	0.3695	0.1447	0.0747	0.0561	0.0470	0.0464

The error rate on the test set using 18 hidden units is 0.0502

2.b

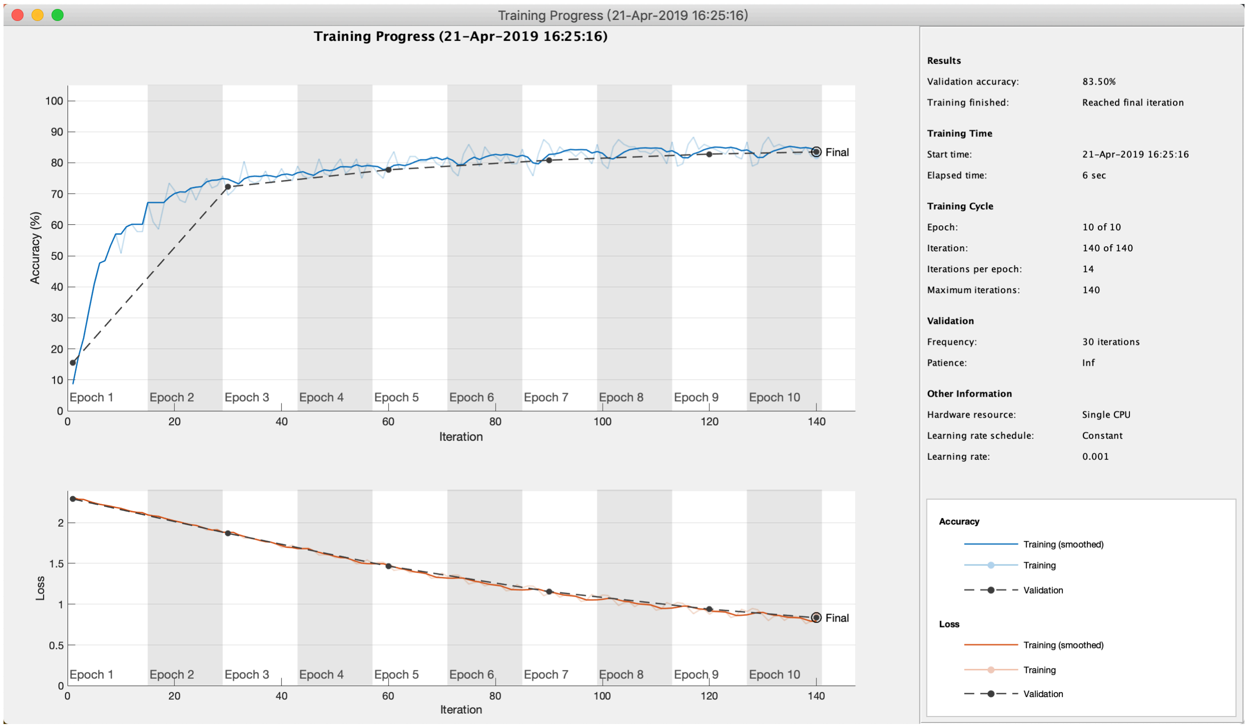




Here I use log(pc+1) scale, 'plus 1' is because we have small negative value in pc. The clustering structure is pretty clear in the 3D plot, while there is lots of overlapping in the 2D plot.

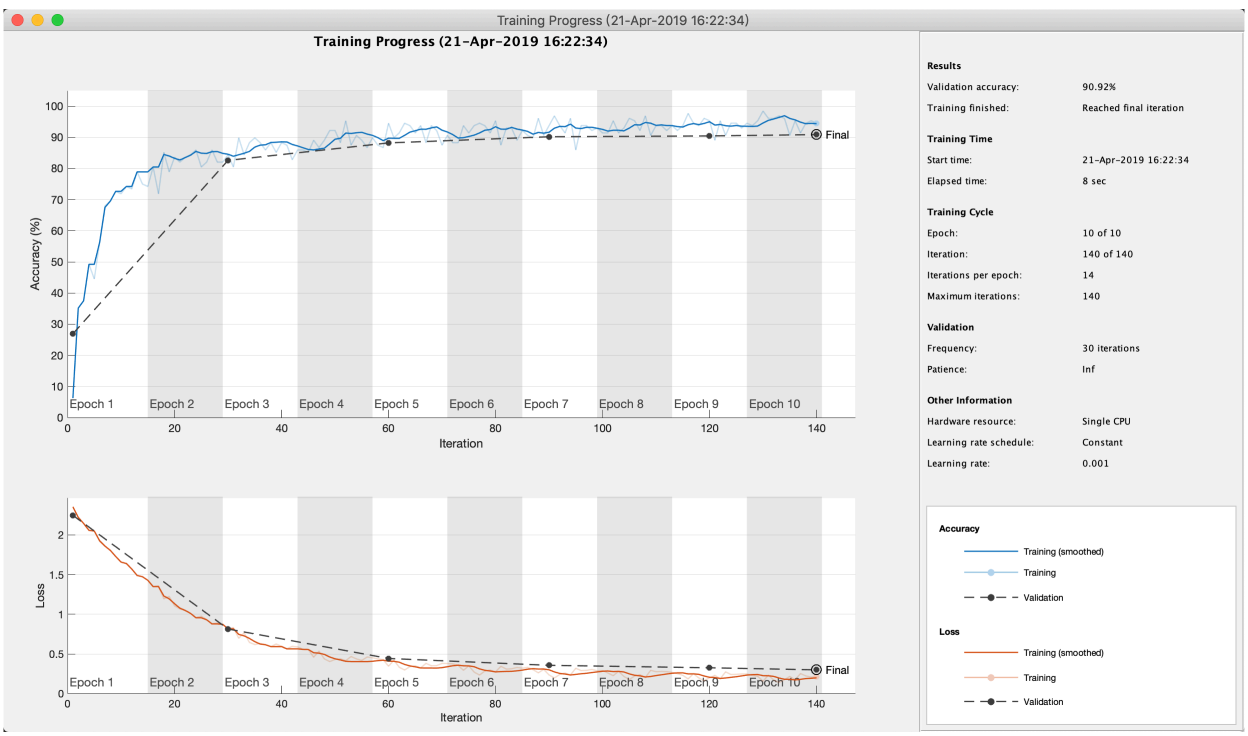
I think this is because that first two components fail to explain the majority of the variances of the data.

3.b



The accuracy on the testing data is 0.8436

3.c



The accuracy on the testing data is 0.9036