(1)

$$E(W, v/x) = -\sum_t \left(r^t \log y^t + (1-r^t)\log(1-y^t)\right) + \sum_h \|w_h\|^2$$

For calculating Δv

$$\frac{\partial E}{\partial v} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial v}$$

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

$$= -\sum_t \frac{r^t - r^t y^t - y^t + y^t r^t}{y^t(1-y^t)}$$

$$= -\sum_t \frac{r^t - y^t}{y^t(1-y^t)}$$

$$\frac{\partial y}{\partial v} = \left(1 - \tanh^2\left(\sum_{h=1}^{H} v_h z_h + v_0\right)\right) \times z_h$$

$$= (1 - y^2) \times z_h$$

$$\frac{\partial E}{\partial v} = \sum_t \frac{r^t - y^t}{y^t (1 - y^t)} (1 - y^2) z_h^t$$

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

$$\Delta v_{ih} = h \sum_t \frac{r_i^t - y_i^t}{y_i^t} (1 + y_i^t) z_h^t$$

$$\frac{\partial E}{\partial w_h}$$

$$= \frac{\partial}{\partial w} - \sum_t (r^t \log y^t + (1 - r^t) \log(1 - y$$

$$+ \text{\$} 2 w_h \quad \text{---} \quad (1)$$

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial z} * \frac{\partial z}{\partial w}$$

$$\frac{\partial y}{\partial z}$$

$$= 1 - \tanh^2\left(\sum_{h=1}^{H} v_h z_h + v_0\right) \times v_h$$

$$= (1 - y^2)\, v_h$$

$$\frac{\partial z}{\partial w} = \begin{cases} 0.01 \times w_h^T x^T < 0 \\ 1 \quad x^+ \quad \text{otherwise} \end{cases}$$

$$= \max(0, w_h^T x^T) + 0.01 \times \min(w_h^T x^T, 0)$$

$$\frac{x^T}{w_h^T x^T}$$

$$= \max(0, w_h^T x^T) + 0.01 \times \min(w_h^T x^T, 0)$$

$$\frac{}{w_h^T}$$

From equation 1

$$= 2w_h + -\sum_t \frac{r^t - y^t}{y^t(1-y^t)} * (1-y^t) \cdot v_h \cdot \max(0, w_h^T x^T) + 0.01 \min(w_h^T x^T, 0)$$

$$= 2w_h * -\sum_t \frac{r^t - y^t}{y^t} (1+y^t) \cdot v_h \cdot \max(0, w_h^T x^T) + 0.01 \min(w_h^T x^T, 0)$$

$$\nabla w_{hj} = h \left(\sum_t \frac{r_i^t - y_i^t}{y^t} (1+y^t) v_h \max(0, w_{hj}^T x_j^T) + 0.01 \min(w_{hj}^T x_j^T, 0) \right)$$

$$- 2 w_{hi}^t$$

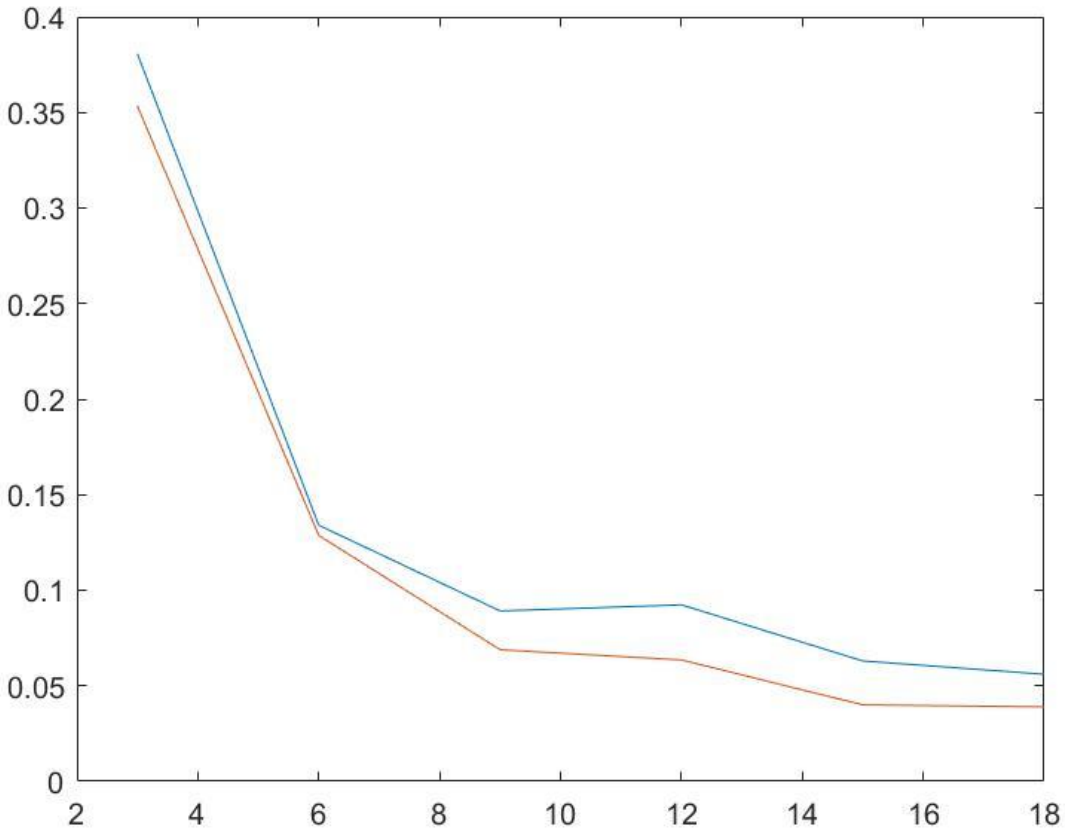
2a

Red :training error

Blue :validation error

Y:error rate

X:num of hidden Unit

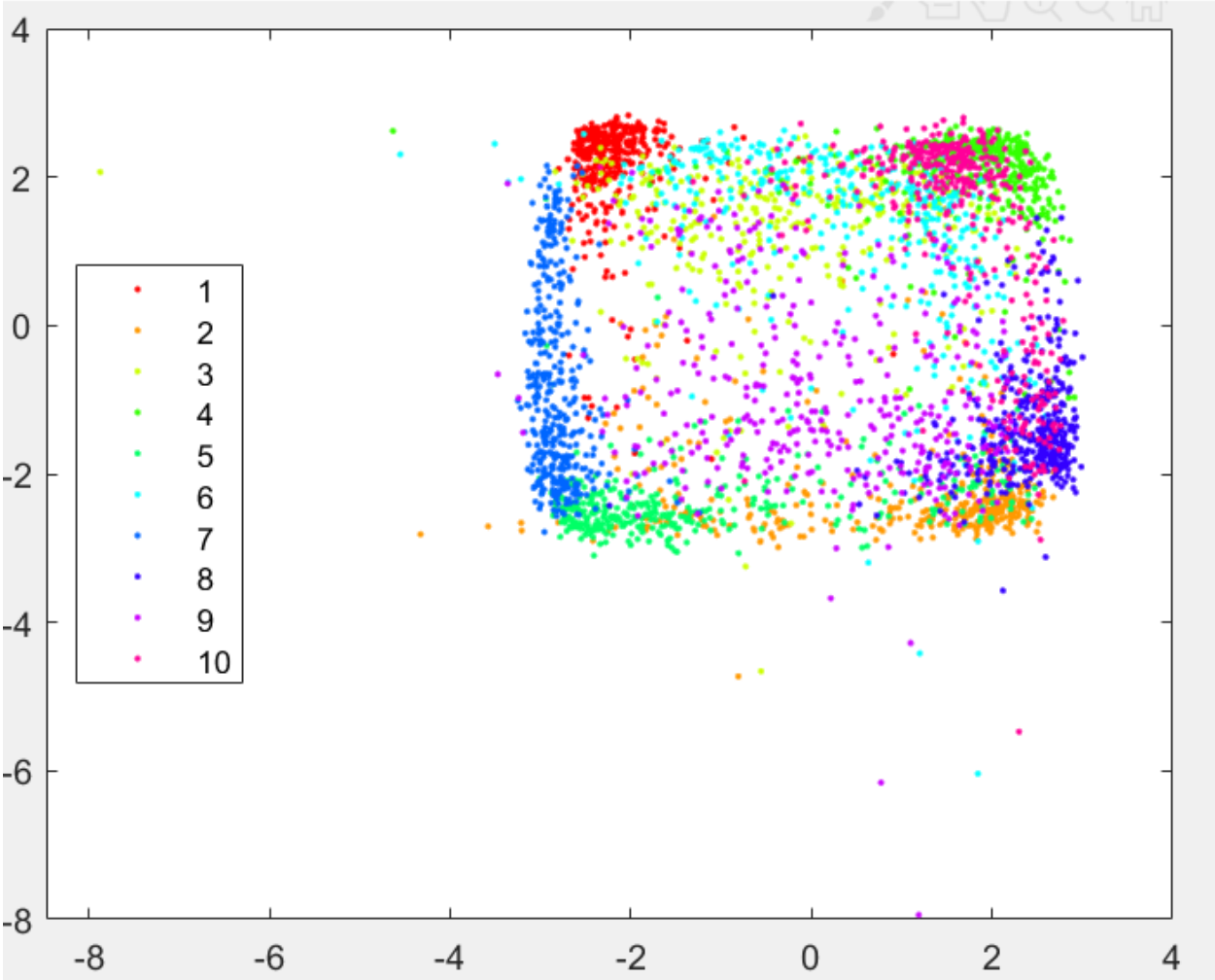


No of Hidden unit	3	6	9	12	15	18
Test data Error Rate	0.3954	0.1361	0.0800	0.0806	0.0576	0.0507
Train error rate	0.35344	0.12867	0.068873	0.063534	0.040043	0.038975
validation error rate	0.38067	0.13401	0.089162	0.092365	0.063001	0.05606

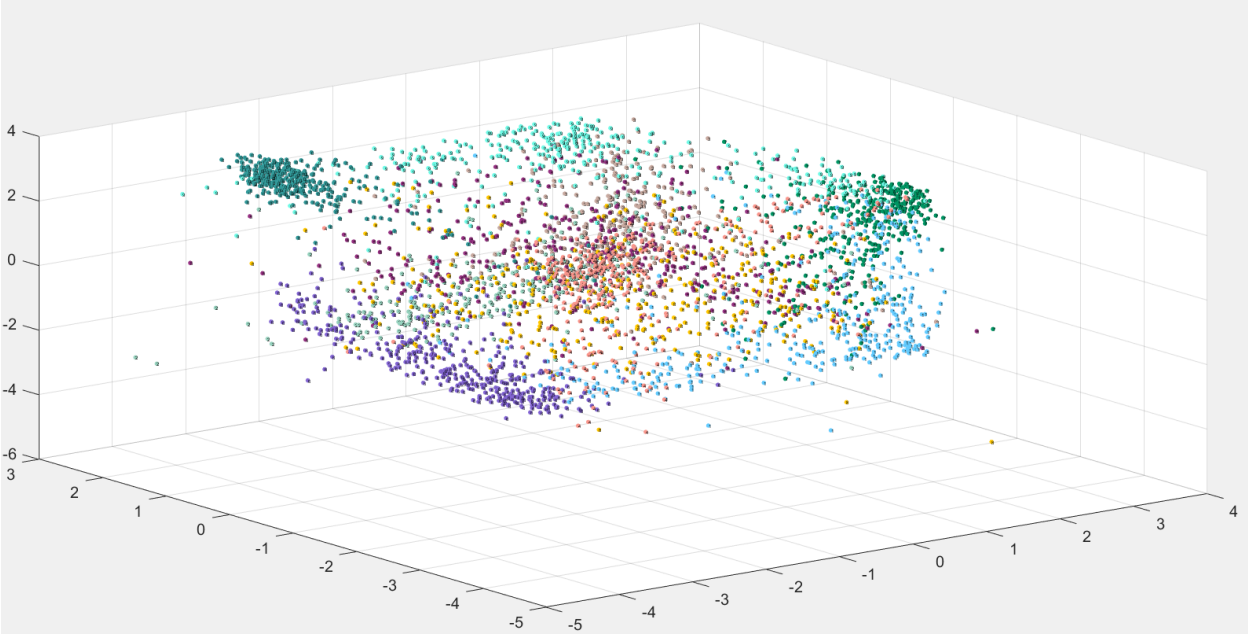
I should use 18 hidden layer since I am getting 0.3417 train error rate, 0.539 validation error rate and 0.0539 test error rate.

2b

2d plot of hidden matrix after Applying PCA and Log



3d plot of hidden layers



Since gscatter for 3d is missing I have printed the name of the cluster on the node itself.

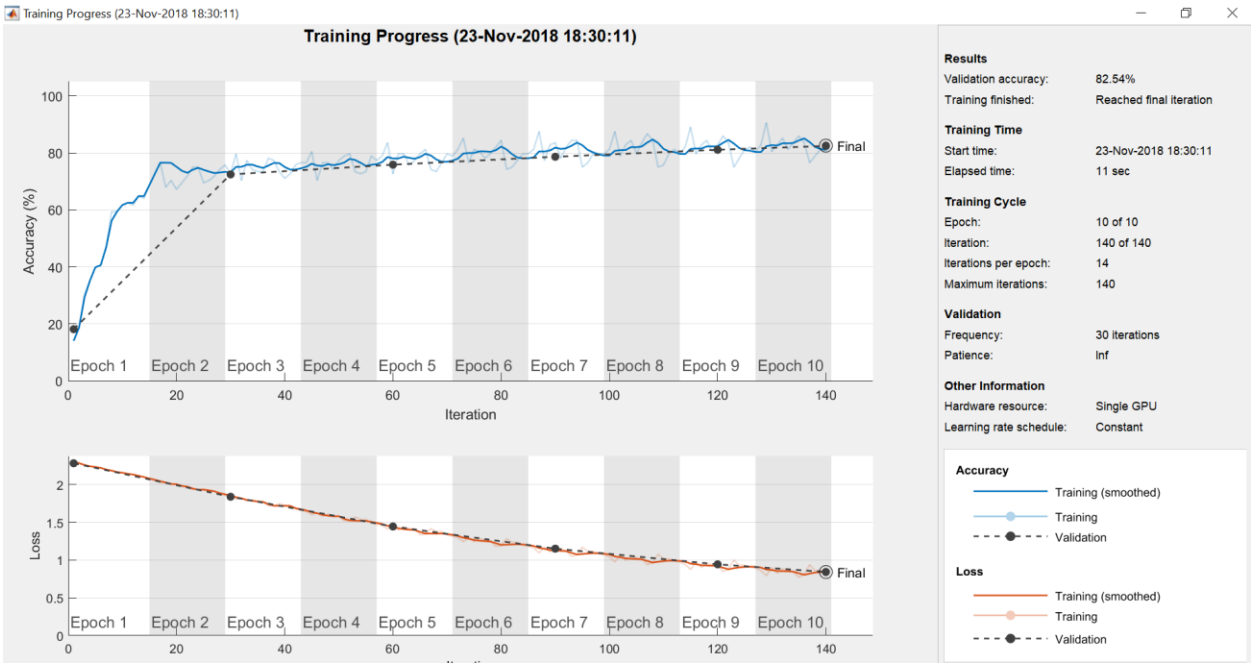
Log method:since some of the values were negative I have converted them to positive then taken the log and then multiplied -1 with it.

In the 3d plot we can visualize each cluster separately but in the 2d plot the clusters were overlapping.

Because of the use of relu function the value will be either zero or its original value and since in the given data most of the points are positive so its coming as its proper value. And some of them are visible -ve since log of its value is -ve.

3 c

accuracy =0.8308



3c ii

accuracy =

0.8997