

## Problem 1.

$$E(w, v | x) = - \sum_t (r^t \log y^t + (1-r^t) \log (1-y^t)) + \sum_h \|w_h\|_2^2$$

$$y^t = \tanh\left(\sum_{h=1}^H v_h z_h^t + w_0\right)$$

$$z_h^t = \text{ReLU}(w_h^t x^t)$$

$$2w_{hj}$$

$$\Delta w_{hj} = -\eta \frac{\partial E}{\partial w_{hj}} = -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_h^t} \frac{\partial z_h^t}{\partial w_{hj}} \quad \left| \quad \frac{\partial \sum_h \|w_h\|_2^2}{\partial y^t} \cdot \frac{\partial y^t}{\partial z_h^t} \cdot \frac{\partial z_h^t}{\partial w_{hj}} = 2w_{hj}$$

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

$$\frac{\partial y^t}{\partial z_h^t} = (1-(y^t)^2) \cdot z_h^t$$

$$\frac{\partial z_h^t}{\partial w_{hj}} = \begin{cases} 0.01 x_j^t & \text{if } w_h^t x^t < 0 \\ x_j^t & \text{otherwise} \end{cases}$$

$$\Delta w_{hj} = -\eta \left( \frac{r^t - y^t}{y^t(1-y^t)} (1-(y^t)^2) z_h^t \cdot 0.01 x_j^t + 2w_{hj} \right) \text{ if } w_h^t x^t < 0$$

$$- \eta \left( \frac{r^t - y^t}{y^t(1-y^t)} (1-(y^t)^2) z_h^t x_j^t + 2w_{hj} \right) \text{ otherwise}$$

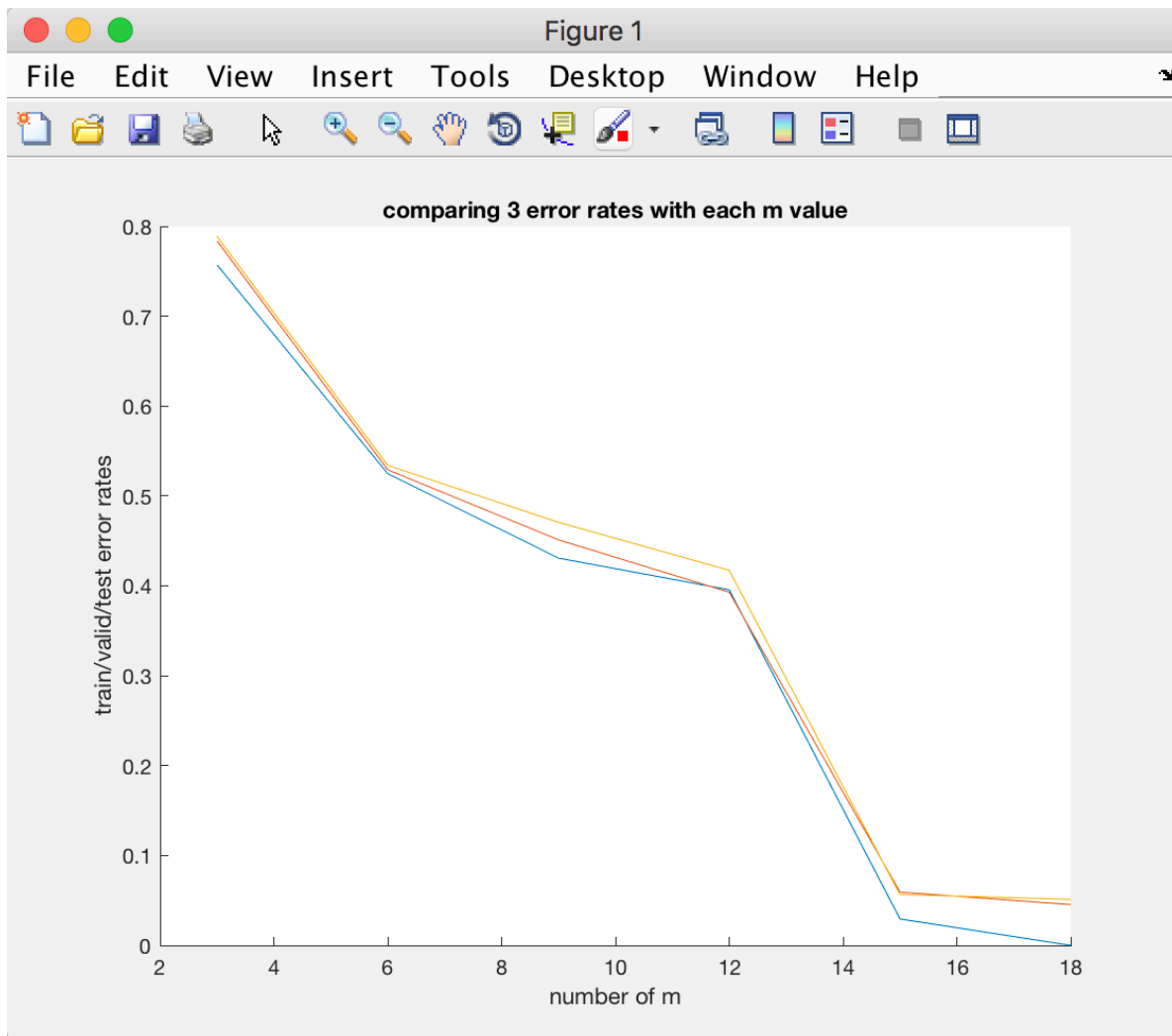
$$\Delta v_h = -\eta \left( \frac{r^t - y^t}{y^t(1-y^t)} (1-(y^t)^2) z_h^t + \frac{2w_h}{0.01 x_j^t} \right) \text{ if } w_h^t x^t < 0$$

$$- \eta \left( \frac{r^t - y^t}{y^t(1-y^t)} (1-(y^t)^2) z_h^t + \frac{2w_h}{x_j^t} \right) \text{ otherwise}$$

Zhiyue Qiu

## Problem2

a. The code runs pretty slow, the data derived is attached.

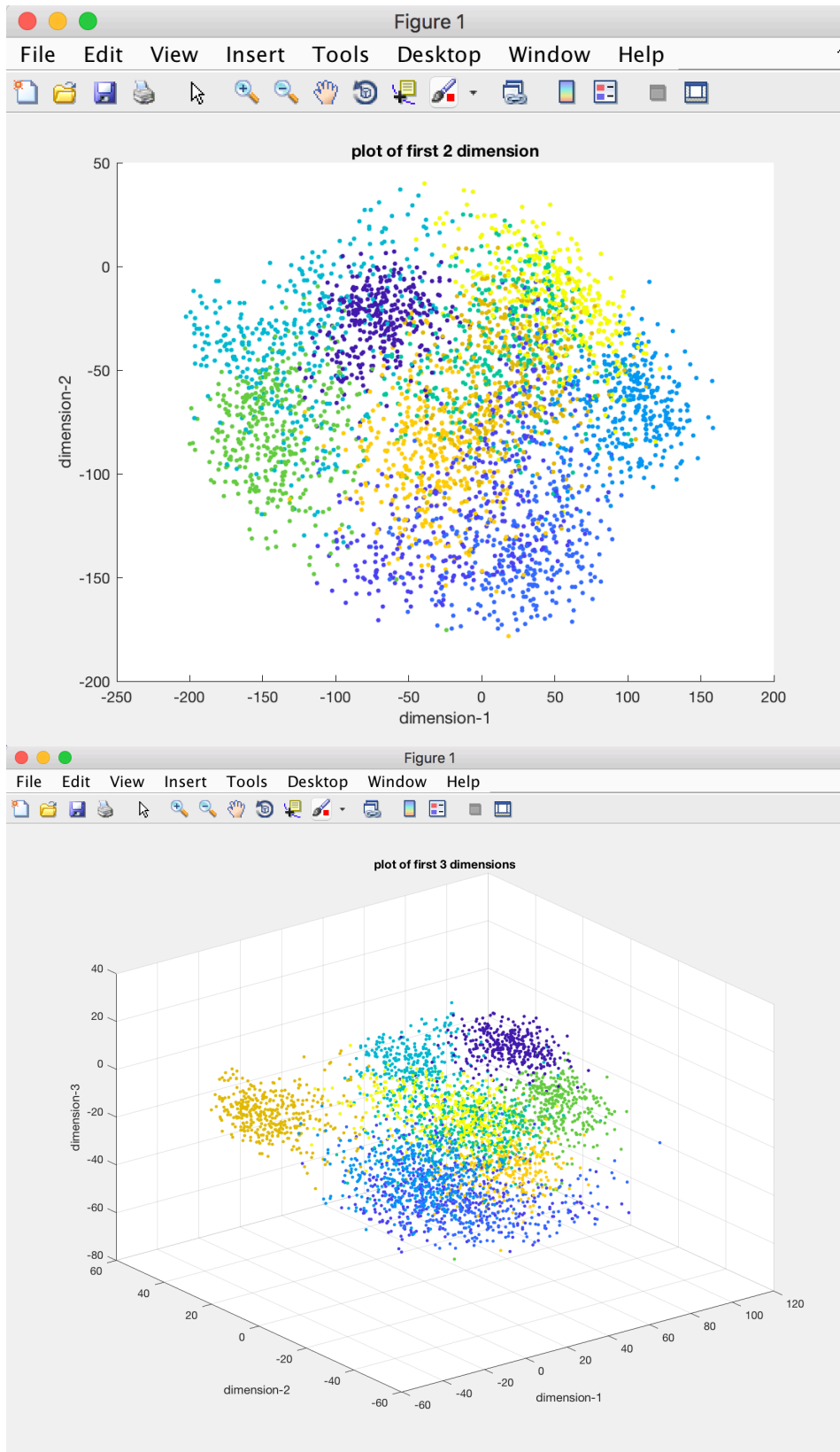


This is the result of three

When  $m = 15$  and  $m = 18$ , I get the best result regarding training, validation and test error rates. But the training error of  $m = 18$  setting is 0, fearing overfitting and  $m = 15$  setting does not significantly worse than  $m = 18$ , so I choose  $m = 15$  in the end.

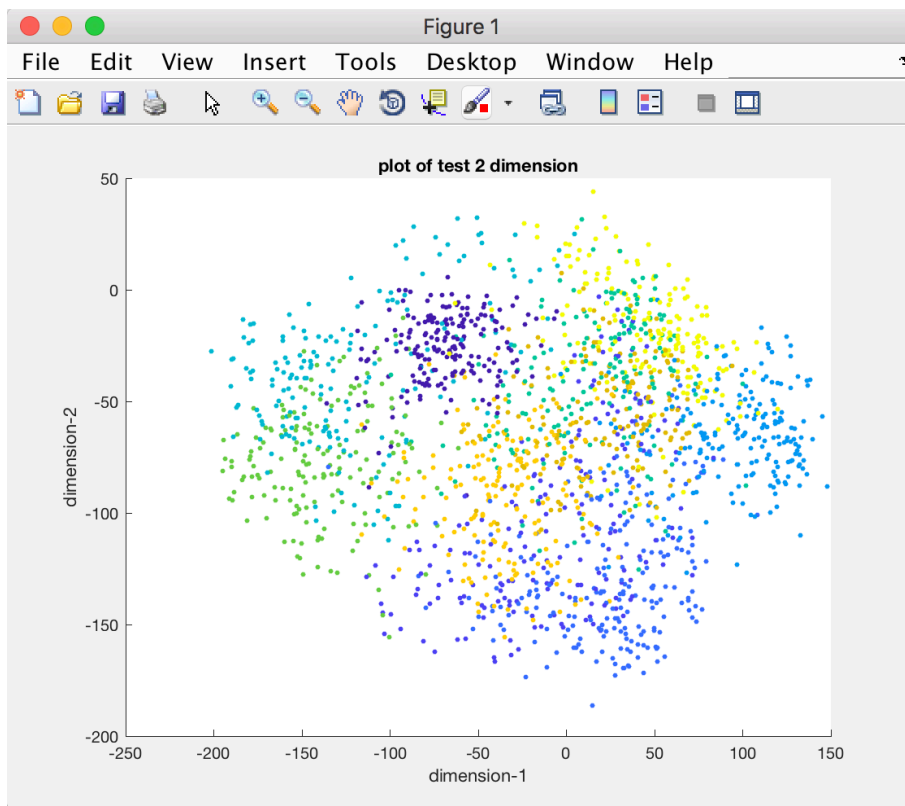
	M = 3	M = 6	M = 9	M = 12	M = 15	M = 18
Training Error	0.757074212493326	0.524826481580352	0.430859583555793	0.395621996796583	0.0293646556326749	0
Validation Error	0.783769353977576	0.529097704217832	0.451147891083823	0.392952482648158	0.0592632140950347	0.0453817405232248
Test Error	0.788687299893277	0.534151547491996	0.470651013874066	0.417289220917823	0.0565635005336179	0.0512273212379936

Problem2b



comparing with 2d and 3d plot, I find it is easier to sperate when we separate them in 3d. That is just visulization, but revealed the data is well separated in high dimension space.

Problem2c



I used the projection vector from previous obtained from training and validation model, and works fine on test data. Data of different tag are separated similar as in training validation data.

