HW 4

Arnab Dey
Student ID: 5563169
Email: dey00011@umn.edu

Solution 1.a

Definitions:

$$
\begin{aligned}
E(\boldsymbol{w_1}, \boldsymbol{w_2}, \boldsymbol{v}|X) &= -\sum_t r^t \ln y^t + (1-r^t)\ln(1-y^t), \\
y^t &= sigmoid\left(\sum_{j=1}^{2}[v_j z_j^t] + v_0\right), \\
z_1^t &= ReLU\left(\sum_{j=1}^{2}[w_{1j} x_j^t] + w_{10}\right), \\
z_2^t &= tanh\left(\sum_{j=1}^{2}[w_{2j} x_j^t] + w_{20}\right), \\
ReLU(x) &= \begin{cases} 0, & \text{for } x < 0, \\ x, & \text{otherwise} \end{cases} \\
ReLU'(x) &= \begin{cases} 0, & \text{for } x < 0, \\ 1, & \text{otherwise.} \end{cases} \\
tanh'(x) &= 1 - tanh^2(x)
\end{aligned} \tag{1}
$$

From the above definitions, we can find the following:

$$
\begin{aligned}
\frac{\partial E}{\partial y^t} &= -\left[\frac{r^t}{y^t} - \frac{1-r^t}{1-y^t}\right] \\
&= -\left[\frac{r^t - y^t}{y^t(1-y^t)}\right], \\
\frac{\partial y^t}{\partial v_j} &= y^t(1-y^t) z_j^t, \quad j \in \{1,2\}, \\
\frac{\partial y^t}{\partial z_j^t} &= y^t(1-y^t) v_j, \quad j \in \{1,2\}, \\
\frac{\partial z_1^t}{\partial w_{1j}} &= ReLU'\left(\sum_{j=1}^{2}[w_{1j} x_j^t] + w_{10}\right) x_j^t, \quad j \in \{1,2\}, \\
\frac{\partial z_2^t}{\partial w_{2j}} &= \left(1 - tanh^2\left(\sum_{j=1}^{2}[w_{2j} x_j^t] + w_{20}\right)\right) x_j^t, \quad j \in \{1,2\}.
\end{aligned} \tag{2}
$$

Based on the above equations we can find the weight updates as follows:

$$
\begin{aligned}
\Delta v_j &= -\eta \frac{\partial E}{\partial v_j}, \quad j \in \{1,2\} \\
&= -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial v_j} \\
&= \eta \sum_t [r^t - y^t] z_j^t \\
\Delta v_0 &= -\eta \frac{\partial E}{\partial v_0} \\
&= -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial v_0} \\
&= \eta \sum_t [r^t - y^t]
\end{aligned} \tag{3}
$$

and,

$$
\begin{aligned}
\Delta w_{1j} &= -\eta \frac{\partial E}{\partial w_{1j}}, \quad j \in \{1,2\} \\
&= -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_1^t} \frac{\partial z_1^t}{\partial w_{1j}} \\
&= \eta \sum_t \left[(r^t - y^t) v_1 ReLU' \left(\sum_{j=1}^{2} [w_{1j} x_j^t] + w_{10} \right) x_j^t \right] \\
\Delta w_{2j} &= -\eta \frac{\partial E}{\partial w_{2j}}, \quad j \in \{1,2\} \\
&= -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_2^t} \frac{\partial z_2^t}{\partial w_{2j}} \\
&= \eta \sum_t \left[(r^t - y^t) v_2 \left(1 - tanh^2 \left(\sum_{j=1}^{2} [w_{2j} x_j^t] + w_{20} \right) \right) x_j^t \right] \\
\Delta w_{10} &= -\eta \frac{\partial E}{\partial w_{10}} \\
&= -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_1^t} \frac{\partial z_1^t}{\partial w_{10}} \\
&= \eta \sum_t \left[(r^t - y^t) v_1 ReLU' \left(\sum_{j=1}^{2} [w_{1j} x_j^t] + w_{10} \right) \right] \\
\Delta w_{20} &= -\eta \frac{\partial E}{\partial w_{20}} \\
&= -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_2^t} \frac{\partial z_2^t}{\partial w_{20}} \\
&= \eta \sum_t \left[(r^t - y^t) v_2 \left(1 - tanh^2 \left(\sum_{j=1}^{2} [w_{2j} x_j^t] + w_{20} \right) \right) \right]
\end{aligned} \tag{4}
$$

Solution 1.b

For shared weights $\boldsymbol{w} = \boldsymbol{w_1} = \boldsymbol{w_2}$, we take the average over corresponding weights from a particular input. This does not change the update equation of $\boldsymbol{v}$. Therefore,

$$\begin{aligned}
\Delta v_j &= \eta \sum_t [r^t - y^t] z_j^t, \qquad j \in \{1, 2\} \\
\Delta v_0 &= \eta \sum_t [r^t - y^t] \\
\Delta w_0 &= \frac{1}{2}(\Delta w_{10} + \Delta w_{20}) \\
\Delta w_1 &= \frac{1}{2}(\Delta w_{11} + \Delta w_{21}) \\
\Delta w_2 &= \frac{1}{2}(\Delta w_{12} + \Delta w_{22}),
\end{aligned} \tag{5}$$

where $\Delta w_{11}, \Delta w_{12}, \Delta w_{21}, \Delta w_{22}$ are found from (4).

Solution 2.a

Error rate on training and validation dataset for different hidden units are shown in the following table: The plot is shown Fig. 1

Hidden units	3	6	9	12	15	18
Training error(%)	15.91	1.38	0.21	0.05	0.05	0.05
Validation error(%)	19.27	7.21	3.58	3.26	3.15	3.58

Table 1: Q2.a: Error-rate Vs. hidden units table

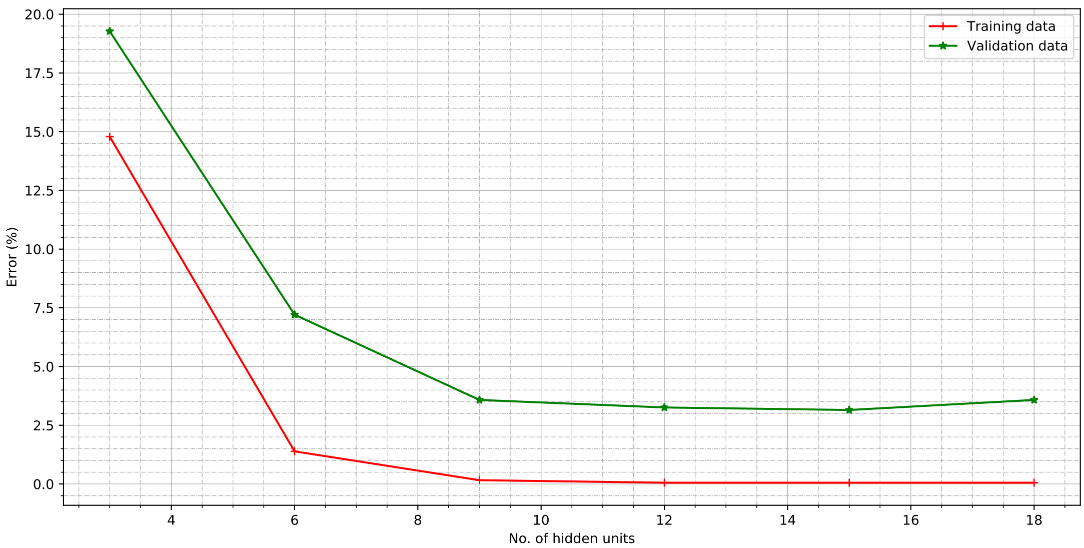

Figure 1: Error Vs. hidden units plot for Optdigits training and validation data

As we can see from Fig. 1, both training and validation error decreases with increase in number of hidden units. But error on validation data with 18 hidden units gets slightly higher compared to 15 hidden units. This implies over-trained MLP. Therefore, for best results, we should choose 15 hidden units.

Error rate on test data with 15 hidden units is shown in the following table:

Hidden units	15
Test error(%)	4.1

Table 2: Q2.a: Error-rate on test data

Solution 2.b

Hidden units used = 15. Hidden representation with 2 principal components is shown in Fig. 2, and with 3 principal components is shown in Fig. 3. Thus from Fig. 2 and Fig. 3, we can see that PCA with 3

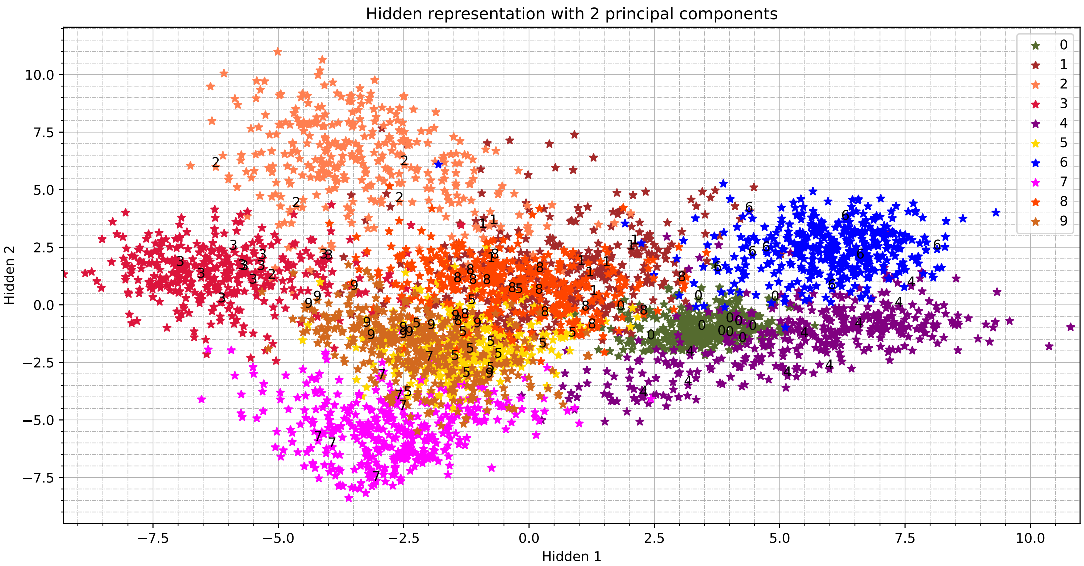

Figure 2: Hidden representation with 2 principal components on training and validation data

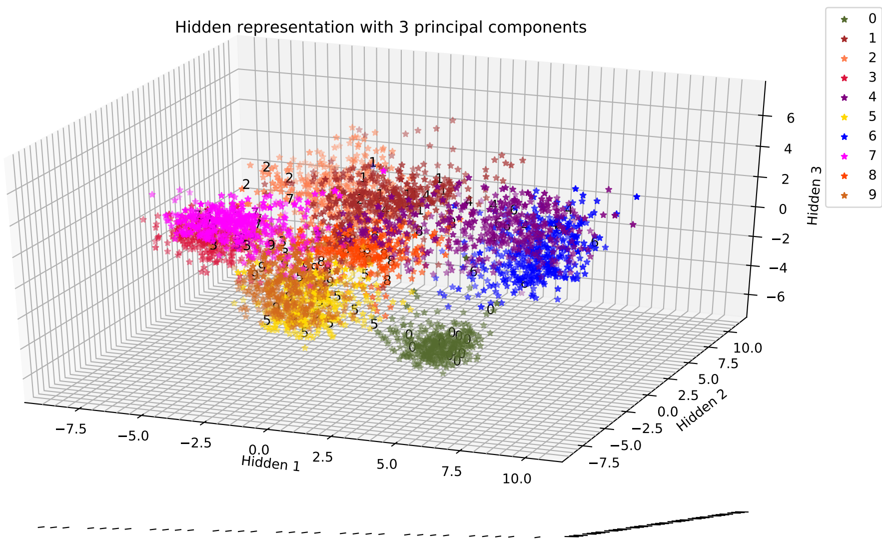

Figure 3: Hidden representation with 3 principal components on training and validation data

principal components provide better separation than with 2 principal components. This happens as number of PCA components behaves as a representation of number of units in the hidden layer.

Solution 3.c

Model 1: Evolution of accuracy with number of epochs is shown in Fig. 4 Evolution of loss with number

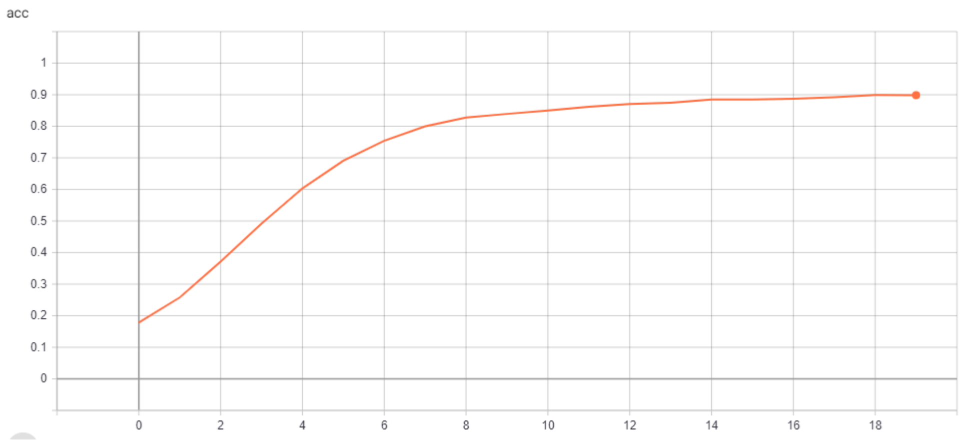

Figure 4: Model 1: Accuracy Vs. Epoch number

of epochs is shown in Fig. 5

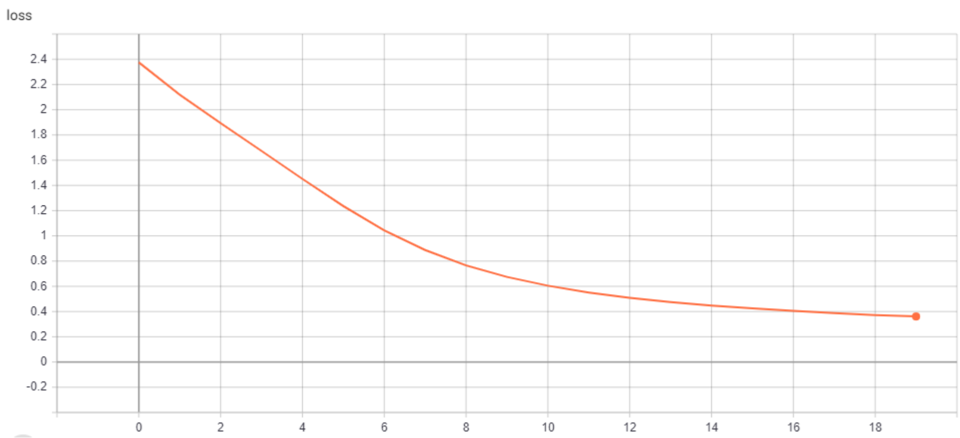

Figure 5: Model 1: Loss Vs. Epoch number

Model 2: Evolution of accuracy with number of epochs is shown in Fig. 6 Evolution of loss with number of epochs is shown in Fig. 7

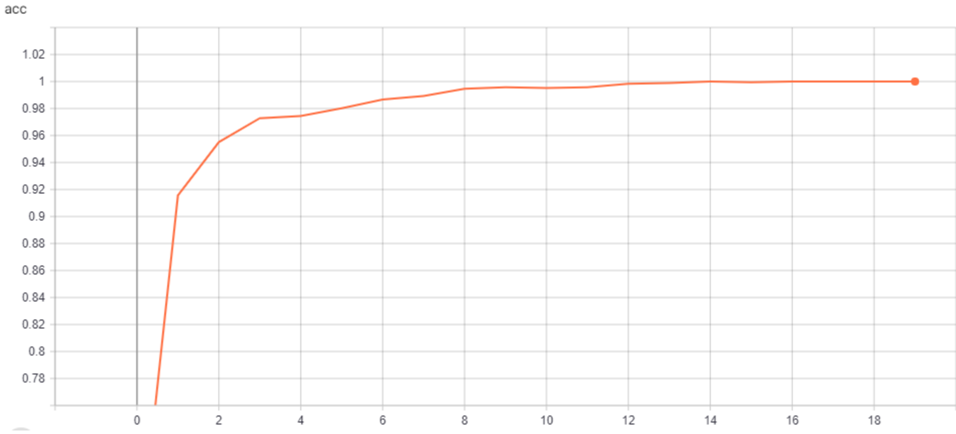

Figure 6: Model 2: Accuracy Vs. Epoch number

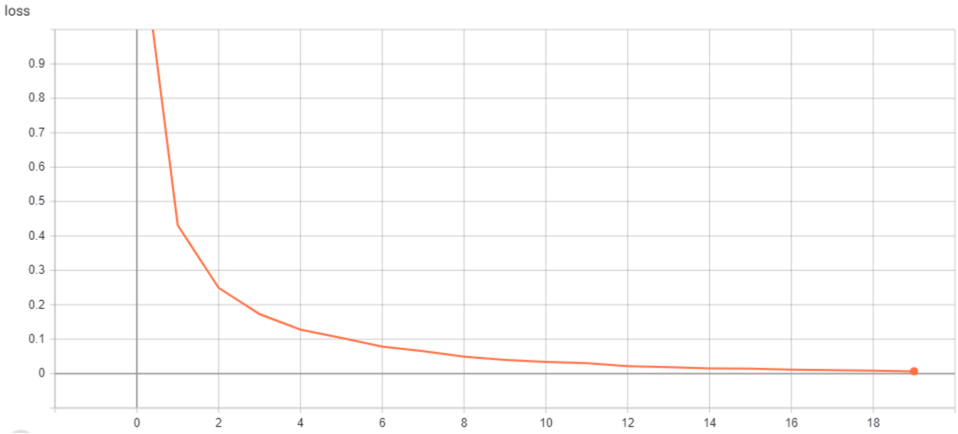

Figure 7: Model 2: Loss Vs. Epoch number

Following table captures the accuracy on test data for both the models:

Model Number	Accuracy(%)	Log file
Model 1	89.1	log_model_1_3c.txt
Model 2	97.5	log_model_2_3c.txt

Table 3: Q3.c: Accuracy on test data