Assignment 1 Theory Problem Set

DO NOT TAG

Name: Jing Yu
GT Email: jyu497@gatech.edu

Theory PS Q1.

Q1 : $S_i = \frac{e^{z_i}}{\sum_k e^{z_k}}$

$\because \frac{\partial}{\partial z_j} \log(S_i) = \frac{1}{S_i} \cdot \frac{\partial S_i}{\partial z_j}$

$\Rightarrow \frac{\partial S_i}{\partial z_j} = S_i \cdot \frac{\partial}{\partial z_j} \log(S_i)$

$= S_i \cdot \left(\frac{\partial z_i}{\partial z_j} - \frac{\partial}{\partial z_j} \log\left(\sum_k e^{z_k}\right)\right)$

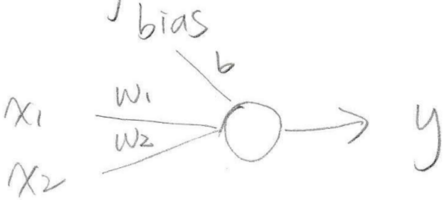
$= S_i \cdot \left(1\{i=j\} - \frac{1}{\sum_k e^{z_k}} \cdot \frac{\partial}{\partial z_j} \sum_k e^{z_k}\right)$

$= S_i \cdot \left(1\{i=j\} - \frac{e^{z_j}}{\sum_k e^{z_k}}\right)$

$= S_i \cdot \left(1\{i=j\} - S_j\right)$

Theory PS Q2.

Q_2 : a single linear threshold neuron

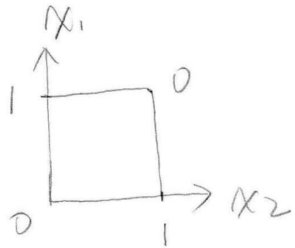


$\Rightarrow$ $W_{AND} = 0.5, 0.5$ $\quad b_{AND} = -0.75$

$W_{OR} = 0.5, 0.5$ $\quad b_{OR} = -0.25$

Theory PS Q3.

Q3 :



You cannot draw a line to separate XOR's plane

$\Rightarrow$ XOR cannot be represented using a linear model

Assignment 1 Writeup

DO NOT TAG

Name: Jing Yu
GT Email: jyu497@gatech.edu

Two-Layer Neural Network

DO NOT TAG

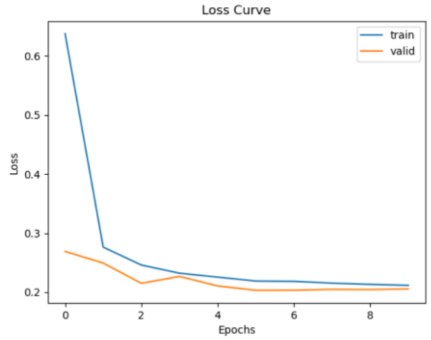
1. Learning Rates

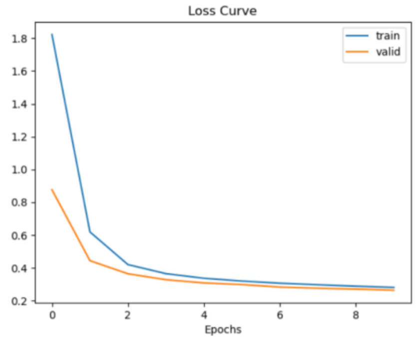
Tune the learning rate of the model with all other default hyper-parameters fixed.

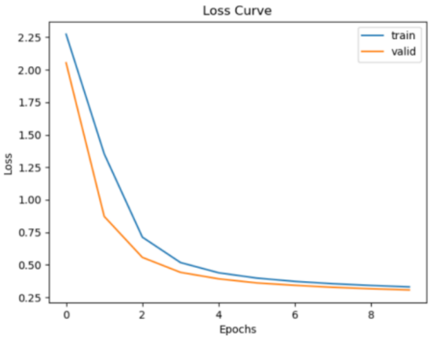
	lr=1	lr=1e-1	lr=5e-2	lr=1e-2
Training Accuracy	0.9426	0.9212	0.9091	0.7293
Validation Accuracy	0.9472	0.9266	0.9156	0.7768
Test Accuracy	0.9441	0.9267	0.9140	0.7638

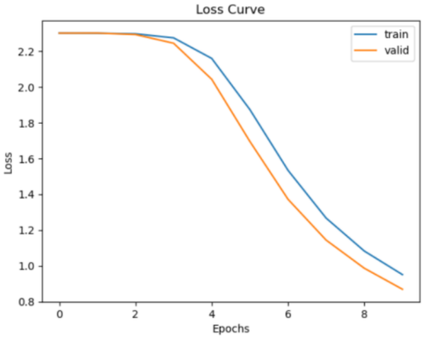
1. Learning Curve

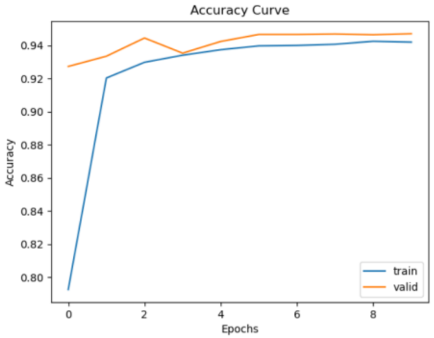
1 0.1 0.05 0.01

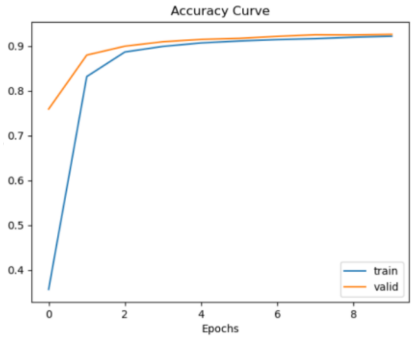


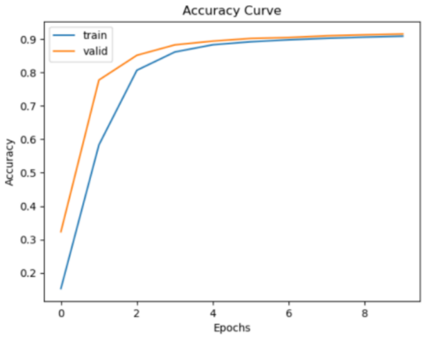


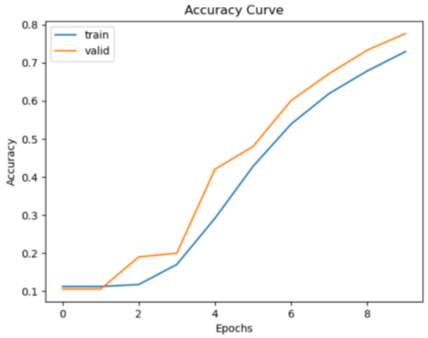












1. Learning Rates

Learning rate controls the amount the weights been updated on each epoch, therefore affects the convergence speed of the model. With a smaller learning rate, the model converge much slower and need more epochs to reach an optimal. Besides, a learning rate that is too small can cause the process to get stuck. Therefore, learning_rate=1 achieved better accuracy given the same epoch value.

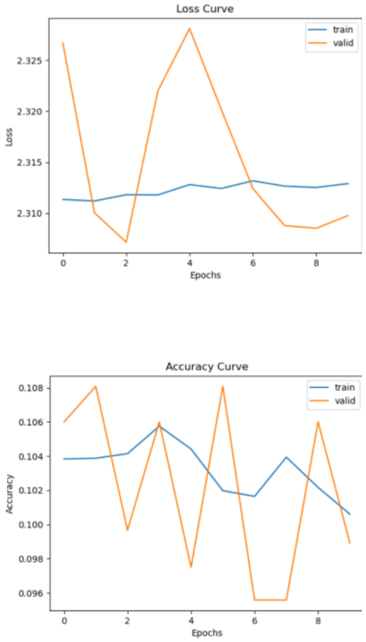
2. Regularization

Tune the regularization coefficient of the model with all other default hyper-parameters fixed.

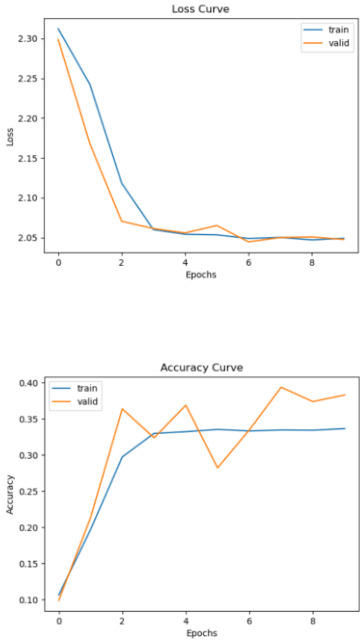
	alpha=1	alpha=1e-1	alpha=1e-2	alpha=1e-3	alpha=1e-4
Training Accuracy	0.1057	0.3364	0.8844	0.9212	0.9295
Validation Accuracy	0.1081	0.3936	0.8961	0.9266	0.9332
Test Accuracy	0.1028	0.3924	0.8958	0.9267	0.9324

2. Regularization

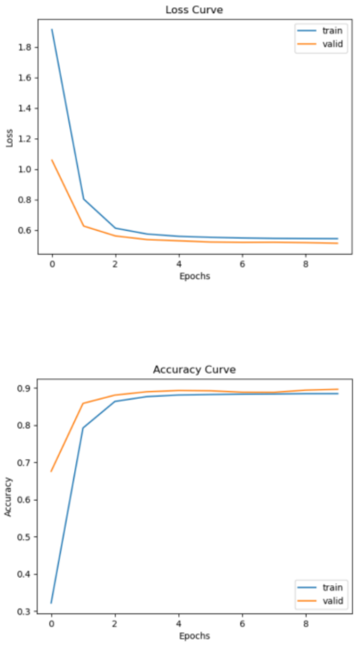
1



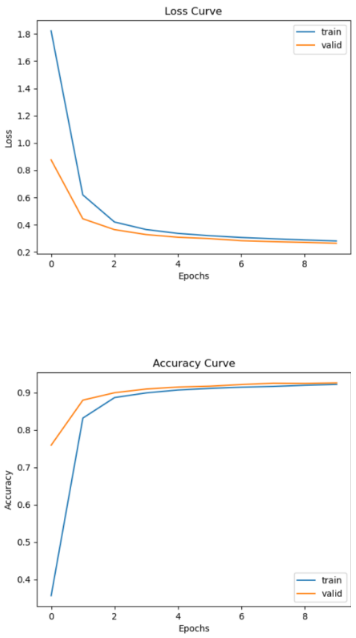
0.1



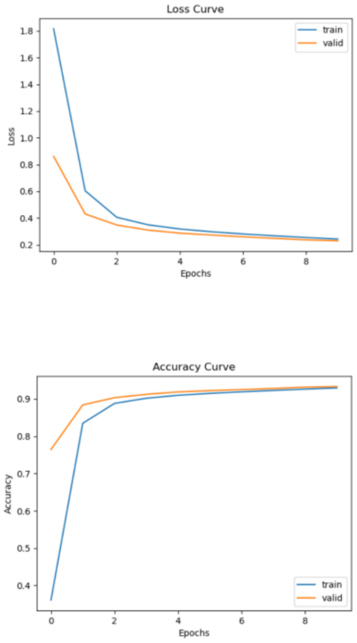
0.01



0.001



0.0001



2. Regularization

The regularization term controls the amount some weights been reduced, therefore controls the complexity of the model. If the alpha is too large, the model won’t learn enough about the dataset and may cause underfitting. If the alpha is too small, the model overcommits to the dataset and may cause overfitting.

The results of alpha=1 and 0.1 show the sign of underfitting. If we further decrease alpha, we may observe overfitting with high training accuracy and lower test accuracy.

3. Hyper-parameter Tuning

Best model and accuracy:

Learning _rate	Regularization	Epochs	Batch _size	Hidden _size	Training Accuracy	Validation Accuracy	Test Accuracy
10	1e-4	15	64	128	0.9458	0.9573	0.9577

The effect of learning rate and regularization term has been discussed in the previous section. With larger epoch value, model will have enough time to converge but it will also increase the training time. Larger batch sizes tend to achieve similar or better training accuracy but lower test accuracy. My intuition is it again overcommit to the dataset. The hidden size will affect the complexity of the model, therefore need to be tuned with the regularization term to achieve best fitting.