

MNIST Digit Classification with MLP

张益铭 车13 2021010552

参数说明

训练参数如下：

```
1 config = {
2     'learning_rate': 0.01,
3     'weight_decay': 1e-4,
4     'momentum': 0.9,
5     'batch_size': 64,
6     'max_epoch': 30,
7     'disp_freq': 100,
8     'test_epoch': 1
9 }
```

一层隐藏层

网络结构：

```
1 model = Network()
2 model.add(Linear('fc1', 784, 128, 0.01))
3 model.add(ActivationFunction('function name'))
4 model.add(Linear('fc2', 128, 10, 0.01))
```

训练数据

训练时间：

	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss
Selu	56s	1m 4s	1m 37s
HardSwish	56s	1m	1m 29s
Tanh	1m 2s	1m 5s	1m 22s

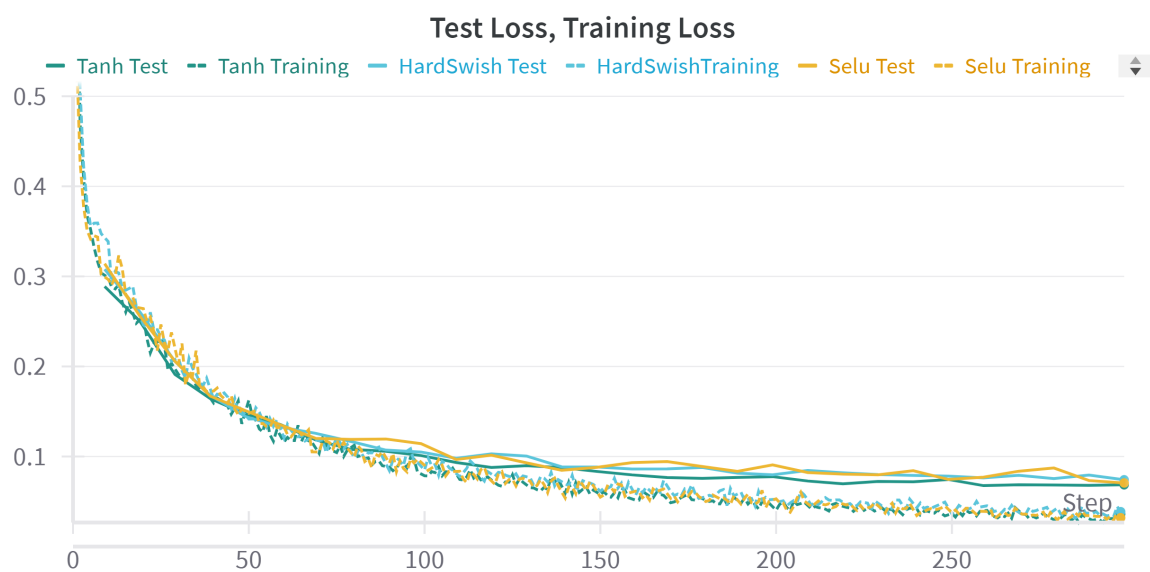
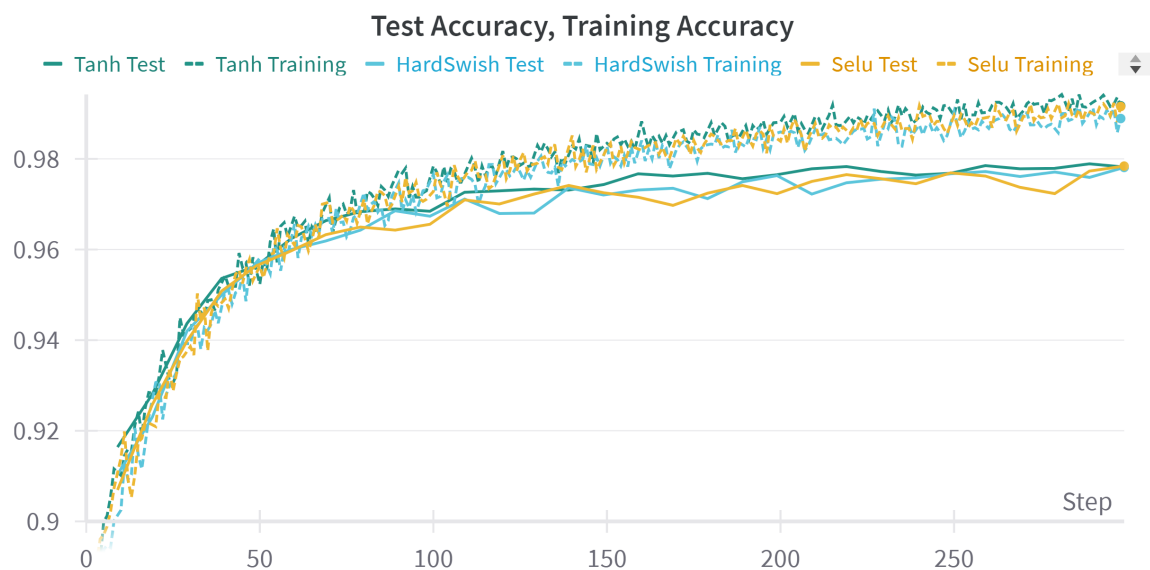
准确率：

	Train			Test		
	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss
Selu	0.99156	0.98906	0.99562	0.9784	0.97383	0.9796
HardSwish	0.98891	0.98969	0.99328	0.97811	0.97462	0.97641
Tanh	0.99172	0.99125	0.99516	0.9782	0.97621	0.97651

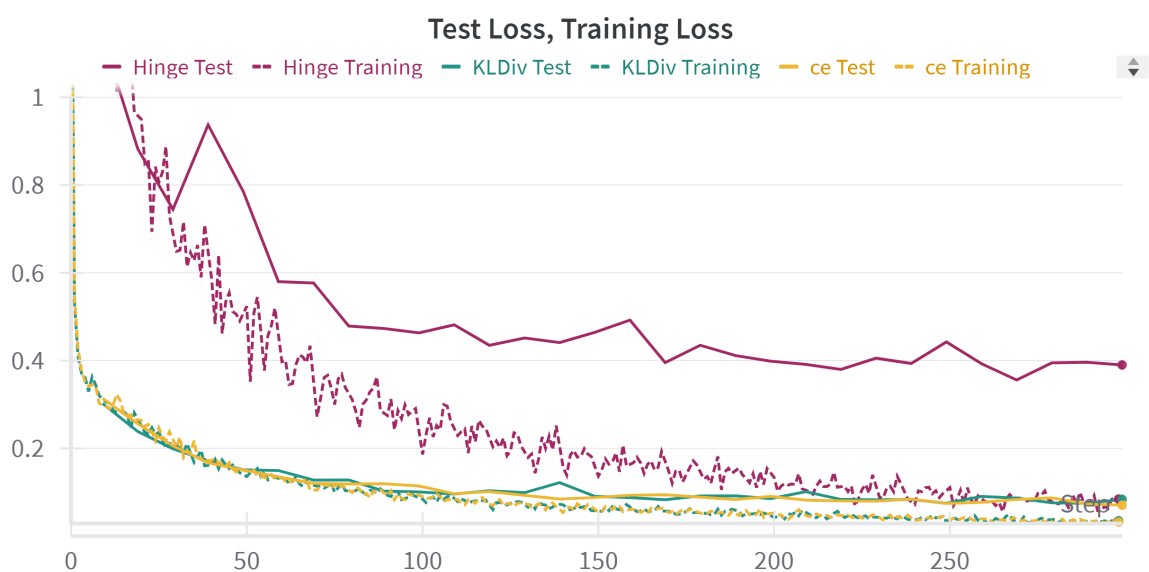
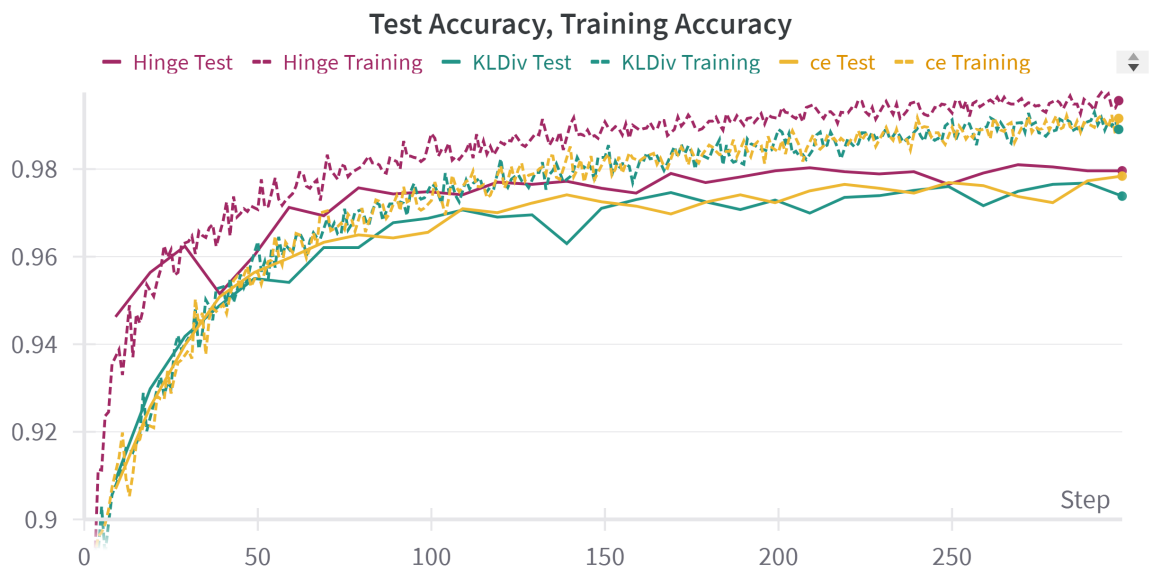
Loss：

	Train			Test		
	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss

	Train			Test		
Selu	0.032006	0.034958	0.084448	0.070628	0.083912	0.3901
HardSwish	0.038739	0.037844	0.09857	0.074165	0.081188	0.45336
Tanh	0.036065	0.035289	0.09165	0.068575	0.073841	0.43273



上图为损失函数为 `SoftmaxCrossEntropyLoss` 的准确率 and loss 曲线



上图为激活函数为 se1u 的准确率和loss曲线

小结

	不同激活函数	不同loss
时间	loss相同时，选用不同激活函数的训练时间差异不大	HingeLoss的训练时间最长，交叉熵损失函数的训练时间最短
收敛速度	不同激活函数对收敛速度的影响较小，说明激活函数并不会影响模型的收敛速度	HingeLoss收敛最慢，而交叉熵和KL散度损失函数都能较快地收敛
准确率	Tanh的准确率整体高于HardSwish和Selu，可能是由于tanh的对称性更好	HingeLoss的准确性明显高于另外两种损失函数；交叉熵和KL散度损失函数准确率大致相同，二者虽然计算复杂度低，但是影响了准确率

两层隐藏层

网络结构：

```
1 model = Network()  
2 model.add(Linear('fc1', 784, 128, 0.01))  
3 model.add(ActivationFunction('function name'))  
4 model.add(Linear('fc2', 128, 64, 0.01))  
5 model.add(ActivationFunction('function name'))  
6 model.add(Linear('fc3', 64, 10, 0.01))
```

训练数据

训练时间：

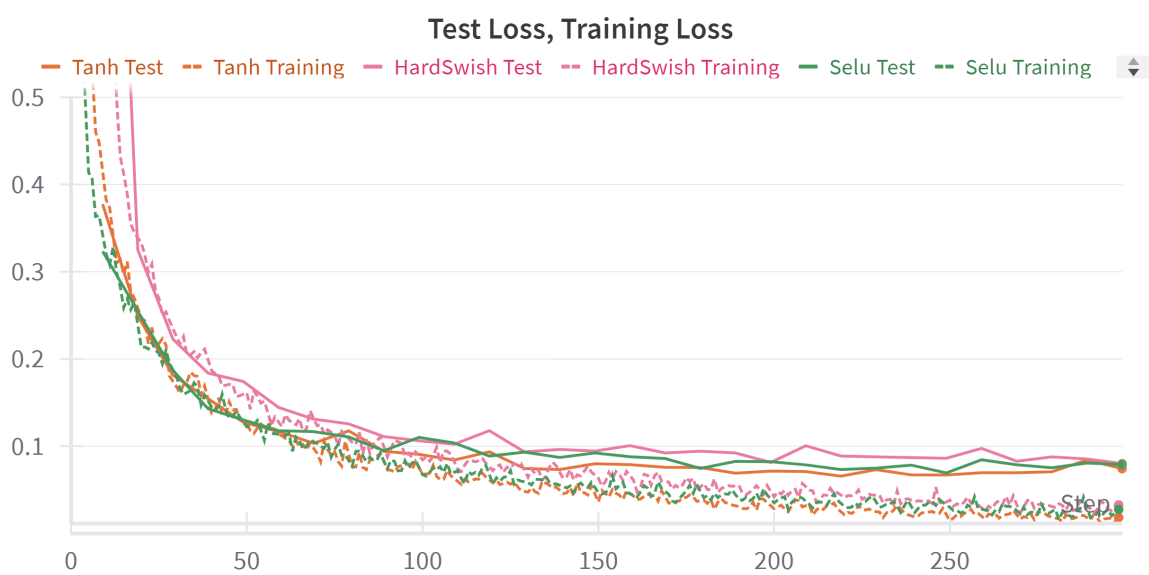
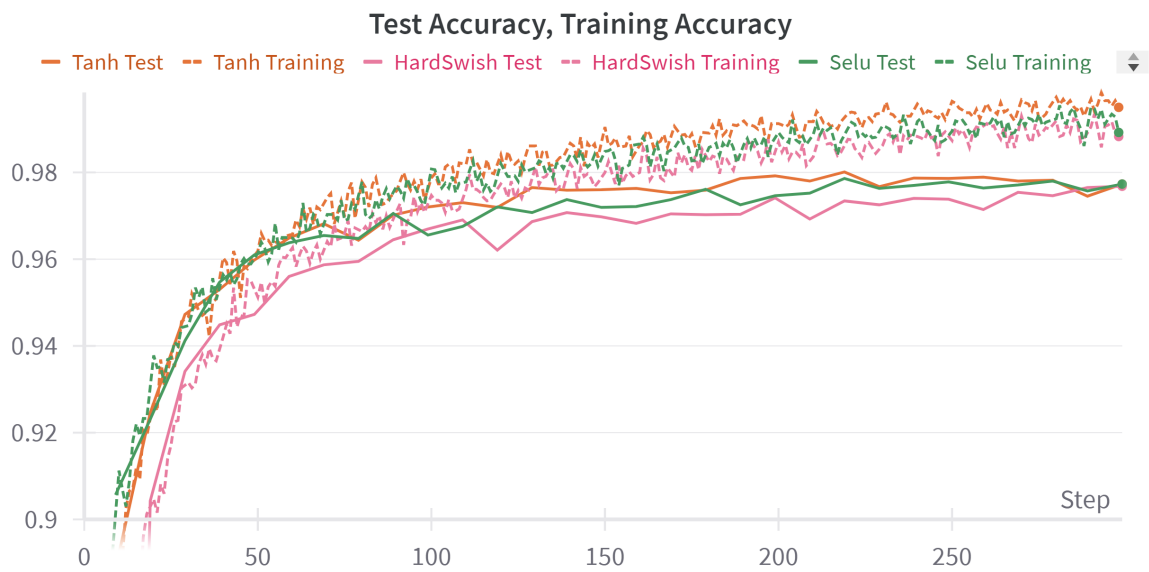
	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss
Selu	1m 3s	55s	1m 31s
HardSwish	1m 5s		
Tanh	1m		

准确率：

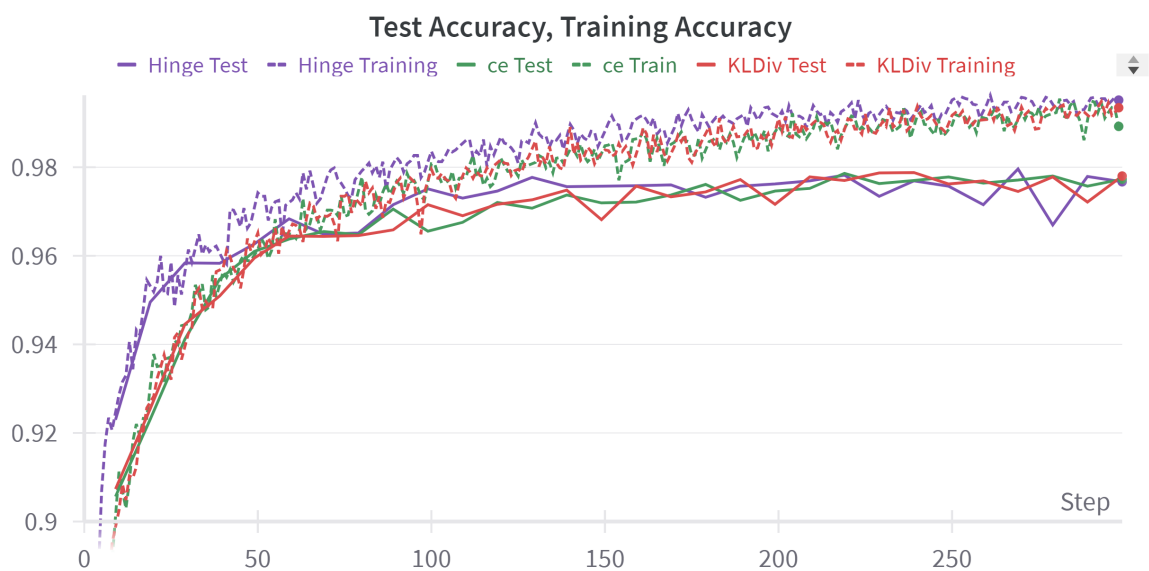
	Train			Test		
	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss
Selu	0.98922	0.99344	0.99516	0.97731	0.97801	0.97671
HardSwish	0.98828			0.97681		
Tanh	0.995			0.97731		

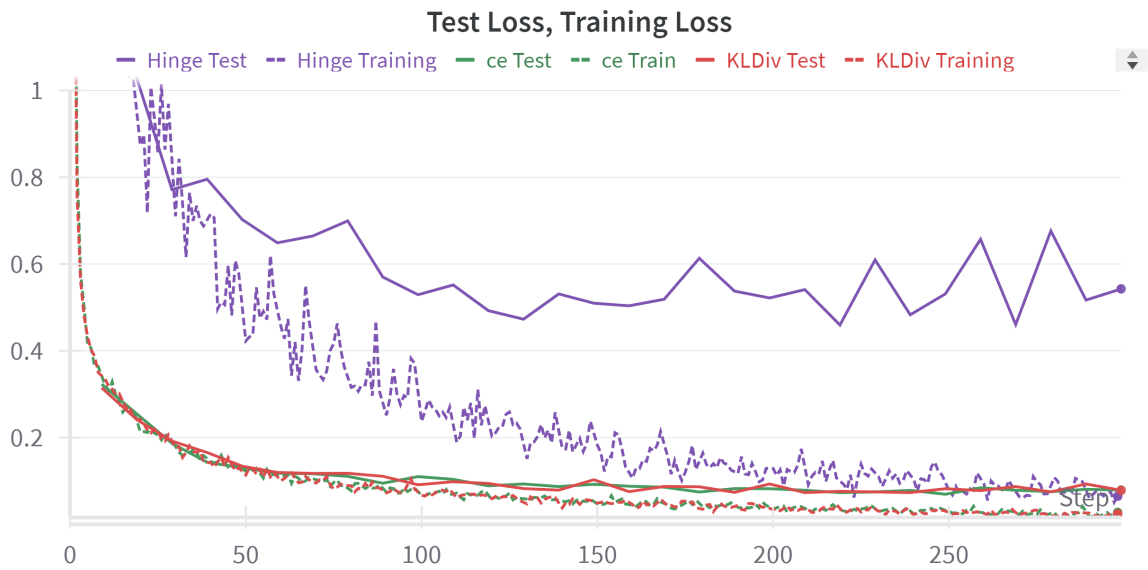
Loss：

	Train			Test		
	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss	SoftmaxCrossEntropyLoss	KLDivLoss	HingeLoss
Selu	0.027503	0.024294	0.064957	0.07897	0.078557	0.54269
HardSwish	0.032726			0.080032		
Tanh	0.018262			0.0743		

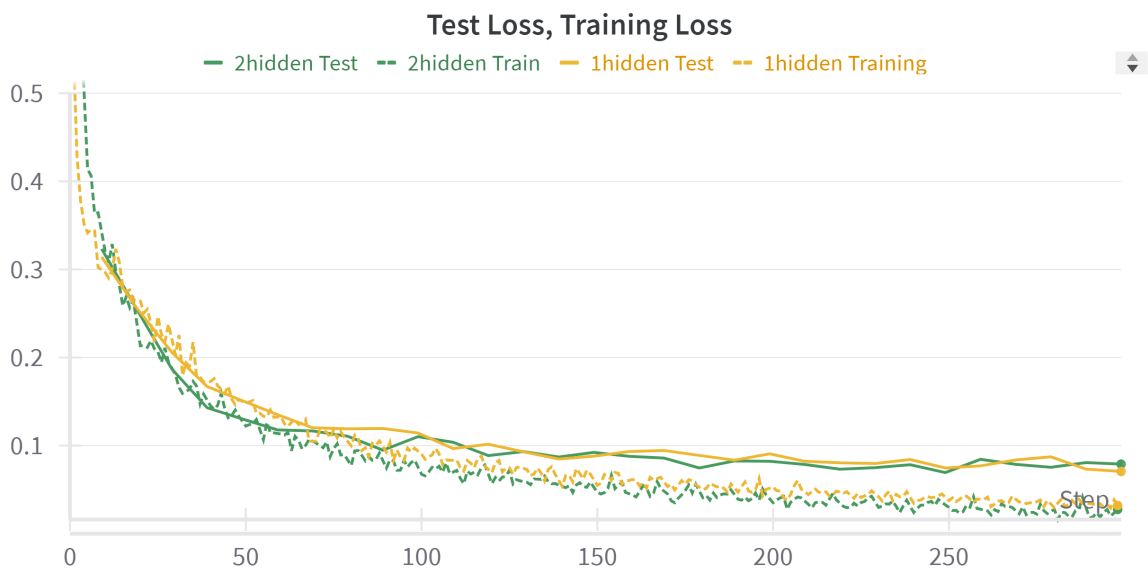
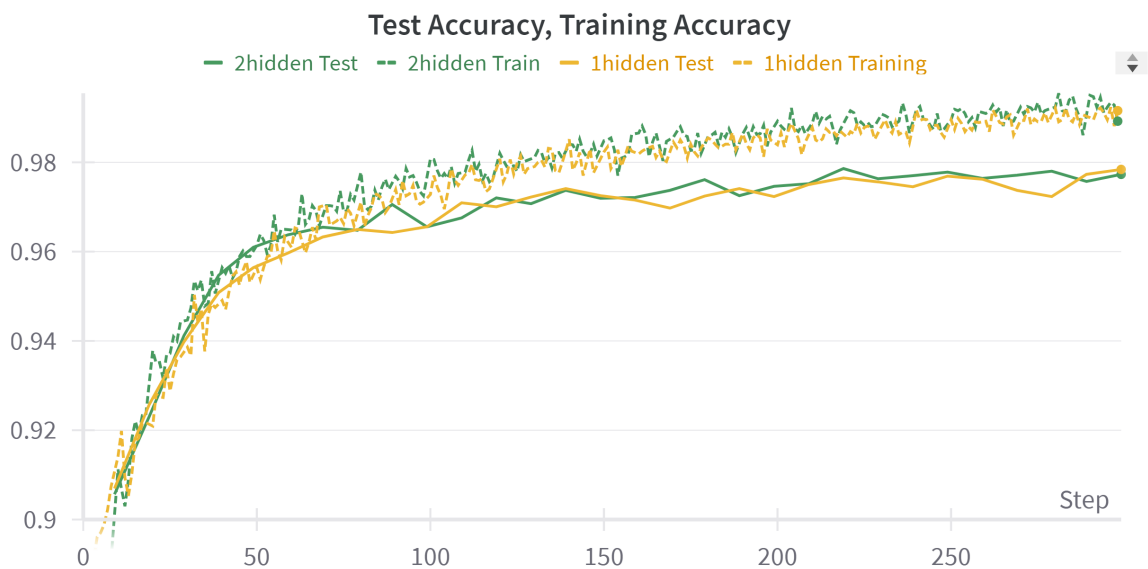


上图为损失函数为 softmaxCrossEntropyLoss 的准确率 and loss 曲线





上图为激活函数为 `se1u` 的准确率和loss曲线



上图为激活函数为 `se1u`，损失函数为 `softmaxCrossEntropyLoss` 的单隐藏层和双隐藏层的准确率和loss曲线

小结

根据上文图片可以看出，由单隐藏层变为双隐藏层时，训练时间、收敛速度、准确率在不同激活函数、不同loss下变化规律几乎相同。

但是当其他条件相同时，增加隐藏层的数目可以一定程度上增加模型的准确率，并提高收敛速度。