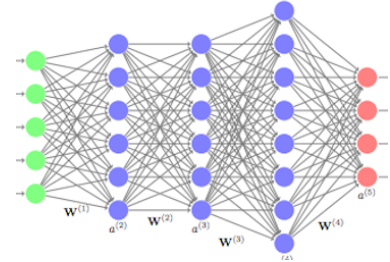


به نام خدا

تمرین سری ۳ یادگیری عمیق - دکتر فاطمی زاده

علی فتحی ۹۴۱۰۹۲۰۵

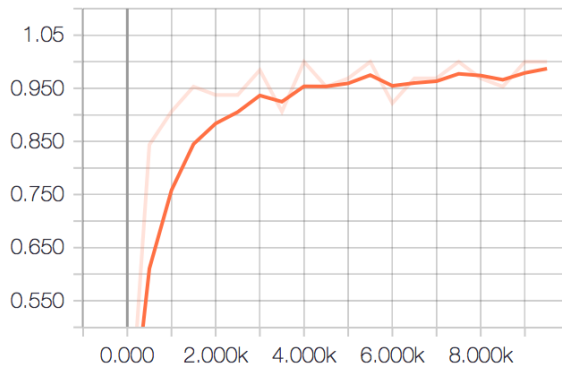


سوال اول:

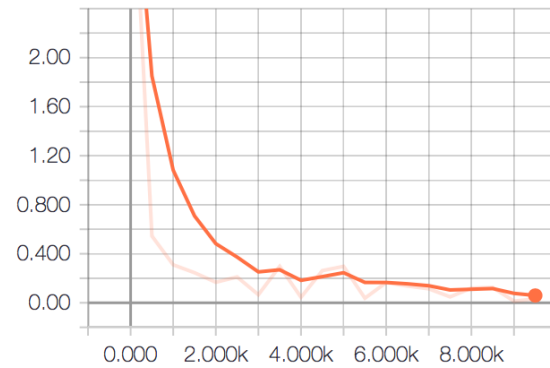
Optimizer: Gradient Descend (Learning Rate = 0.5)
Activation Function: Sigmoid
Hidden Layer: 50 Neurons
Mini-batch size: 64

نتایج به صورت زیر است:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss



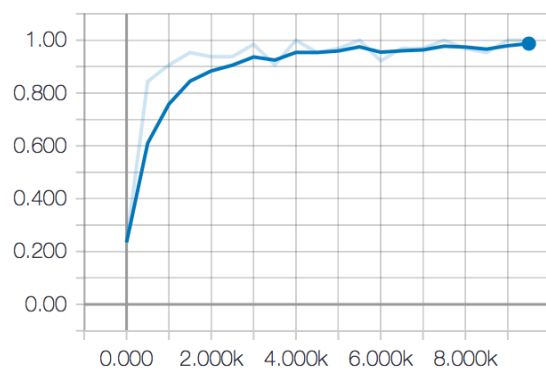
```
step 5000 accuracy is 0.93750 and cross_entropy(loss) is 0.31050
step 5500 accuracy is 0.96875 and cross_entropy(loss) is 0.16640
step 6000 accuracy is 0.93750 and cross_entropy(loss) is 0.15871
step 6500 accuracy is 0.98438 and cross_entropy(loss) is 0.08349
step 7000 accuracy is 0.95312 and cross_entropy(loss) is 0.14668
step 7500 accuracy is 0.96875 and cross_entropy(loss) is 0.09123
step 8000 accuracy is 0.95312 and cross_entropy(loss) is 0.11950
step 8500 accuracy is 0.95312 and cross_entropy(loss) is 0.27292
step 9000 accuracy is 1.00000 and cross_entropy(loss) is 0.03707
step 9500 accuracy is 0.96875 and cross_entropy(loss) is 0.07591
Test Error: 0.98
```

سوال دوم:

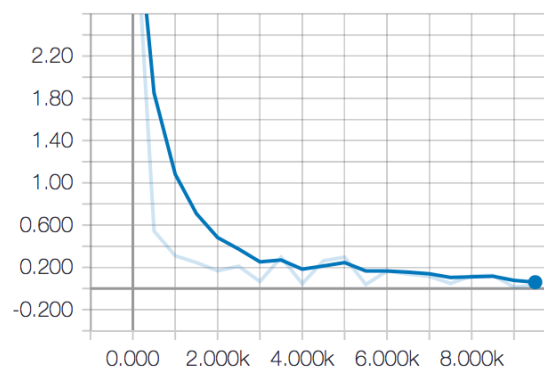
Optimizer: Gradient Descend (Learning Rate = 0.5)
 Activation Function: Tanh
 Hidden Layer: 50 Neurons
 Mini-batch size: 64

نتایج به صورت زیر است:

Accuracy
tag: Save/Accuracy



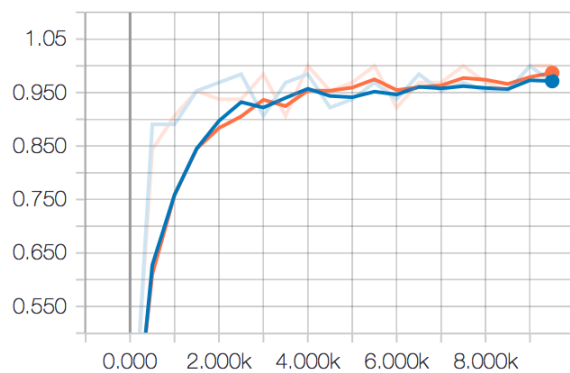
Loss
tag: Save/Loss



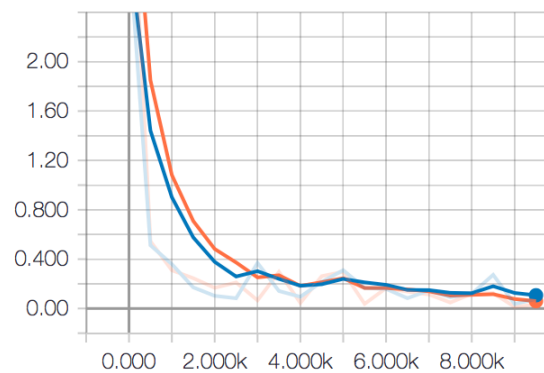
```
step 5000 accuracy is 0.96875 and cross_entropy(loss) is 0.29722
step 5500 accuracy is 1.00000 and cross_entropy(loss) is 0.03762
step 6000 accuracy is 0.92188 and cross_entropy(loss) is 0.16504
step 6500 accuracy is 0.96875 and cross_entropy(loss) is 0.13706
step 7000 accuracy is 0.96875 and cross_entropy(loss) is 0.11401
step 7500 accuracy is 1.00000 and cross_entropy(loss) is 0.04845
step 8000 accuracy is 0.96875 and cross_entropy(loss) is 0.11976
step 8500 accuracy is 0.95312 and cross_entropy(loss) is 0.12486
step 9000 accuracy is 1.00000 and cross_entropy(loss) is 0.01349
step 9500 accuracy is 1.00000 and cross_entropy(loss) is 0.03171
Test Error: 0.92
```

و در مقایسه:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss



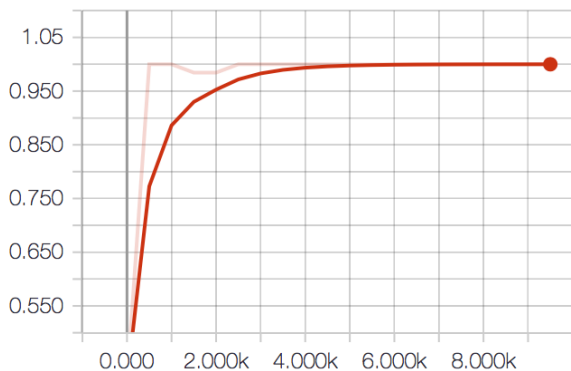
که میبینیم هم در Train، و هم در Test، sigmoid نتیجه بهتری از Tanh دارد.

سوال سوم:

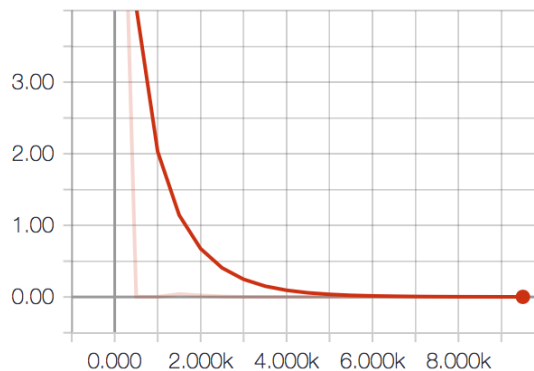
Optimizer: Gradient Descend (Learning Rate = 0.5)
 Activation Function: Tanh
 Hidden Layer: 500 Neurons
 Mini-batch size: 64

نتایج به صورت زیر است:

Accuracy
tag: Save/Accuracy



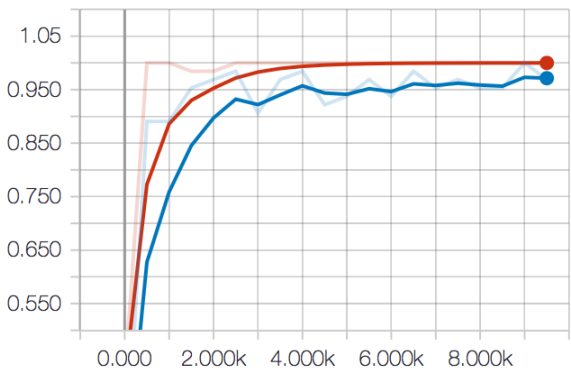
Loss
tag: Save/Loss



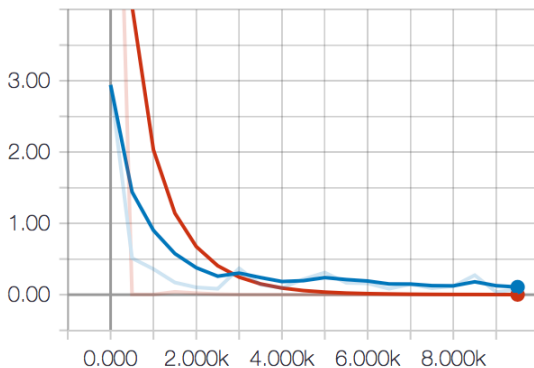
```
step 5000 accuracy is 1.00000 and cross_entropy(loss) is 0.00162
step 5500 accuracy is 1.00000 and cross_entropy(loss) is 0.00061
step 6000 accuracy is 1.00000 and cross_entropy(loss) is 0.00227
step 6500 accuracy is 1.00000 and cross_entropy(loss) is 0.00287
step 7000 accuracy is 1.00000 and cross_entropy(loss) is 0.00070
step 7500 accuracy is 1.00000 and cross_entropy(loss) is 0.00109
step 8000 accuracy is 1.00000 and cross_entropy(loss) is 0.00124
step 8500 accuracy is 1.00000 and cross_entropy(loss) is 0.00174
step 9000 accuracy is 1.00000 and cross_entropy(loss) is 0.00065
step 9500 accuracy is 1.00000 and cross_entropy(loss) is 0.00056
Test Error: 0.98
```

و در مقایسه با Tanh با ۵۰ نورون در لایه مخفی:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss



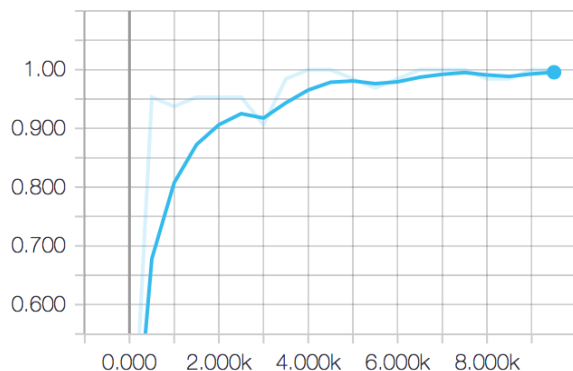
که میبینیم هم در Train، و هم در Test، نتیجه بسیار بهتری از دارد، اما آموزش آن دیرتر همگرا می شود.

سوال چهارم:

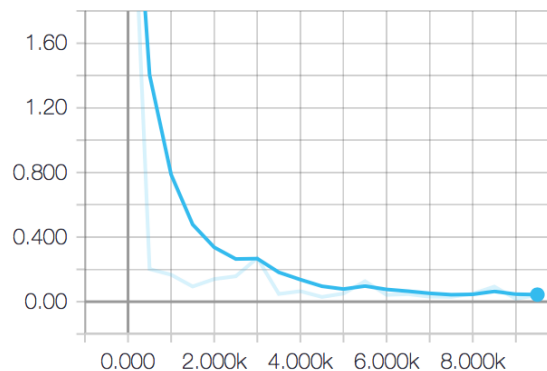
Optimizer: Gradient Descend (Learning Rate = 0.5)
 Activation Function: Tanh
 Hidden Layer 2: 50 Neurons
 Hidden Layer 2: 50 Neurons
 Mini-batch size: 64

نتایج به صورت زیر است

Accuracy
tag: Save/Accuracy



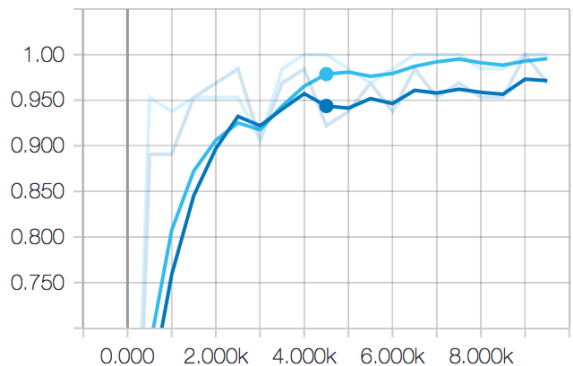
Loss
tag: Save/Loss



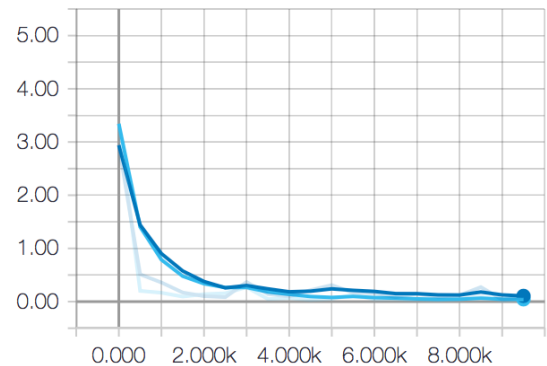
```
step 5000 accuracy is 0.98438 and cross_entropy(loss) is 0.05047
step 5500 accuracy is 0.96875 and cross_entropy(loss) is 0.12672
step 6000 accuracy is 0.98438 and cross_entropy(loss) is 0.04213
step 6500 accuracy is 1.00000 and cross_entropy(loss) is 0.04683
step 7000 accuracy is 1.00000 and cross_entropy(loss) is 0.03083
step 7500 accuracy is 1.00000 and cross_entropy(loss) is 0.02946
step 8000 accuracy is 0.98438 and cross_entropy(loss) is 0.04901
step 8500 accuracy is 0.98438 and cross_entropy(loss) is 0.09326
step 9000 accuracy is 1.00000 and cross_entropy(loss) is 0.01819
step 9500 accuracy is 1.00000 and cross_entropy(loss) is 0.03977
Test Error: 0.96
```

و در مقایسه با Tanh با یک لایه مخفی:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss

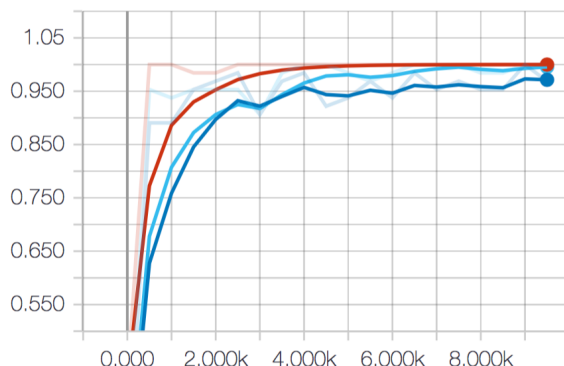


که میبینیم عمیق سازی هم در Train، و هم در Test، نتیجه بهتری از دارد.

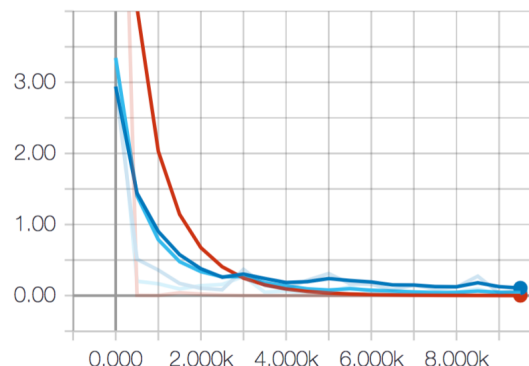
سوال پنجم:

نمودار سوال هر سه سوال ۲ تا ۴ در کنار هم به شکل زیر است:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss



که میبینیم، چه عمیق سازی و چه افزایش تعداد Neuron ها در یک لایه، صحت دسته بندی را بیشتر می کند؛ اما شبکه با دو لایه (منحنی آبی روشن) زمان بیشتری نسبت به شبکه پهن تر (منحنی قرمز) نیاز دارد تا به صحت بالا برسد؛ یعنی آموزش کندتر است.

پارامترهای شبکه عمیق (دو لایه ۵۰ نورونی):

یک اتصال کامل ۷۸۴ به ۵۰ = ۳۹۲۰۰ پارامتر
 یک اتصال کامل ۵۰ به ۵۰ = ۲۵۰۰ پارامتر
 یک اتصال کامل ۵۰ به ۱۰ = ۵۰۰ پارامتر

یعنی شبکه عمیق دو لایه ای در مجموع ۴۲۲۰۰ پارامتر دارد.

پارامترهای شبکه پهن (با ۵۰۰ نورون در لایه مخفی):

یک اتصال کامل ۷۸۴ به ۵۰۰ = ۳۹۲۰۰۰ پارامتر
 یک اتصال کامل ۵۰ به ۱۰ = ۵۰۰ پارامتر

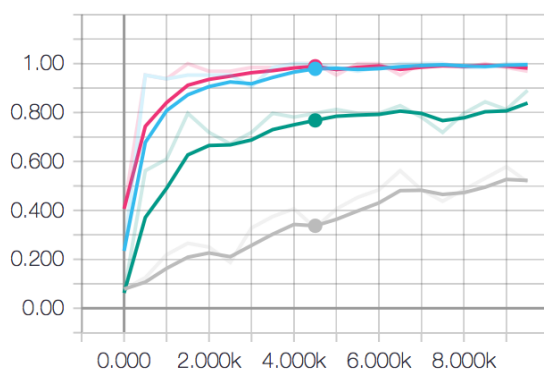
یعنی این شبکه در مجموع ۳۹۲۵۰۰ پارامتر دارد که تقریباً ۱۰ برابر پارامترهای شبکه عمیق است! درحالی که اگر برای آموزش صبر پیشه کنیم شبکه عمیق تا حد قابل قبولی همین کارایی شبکه پهن را خواهد داشت.

سوال ششم:

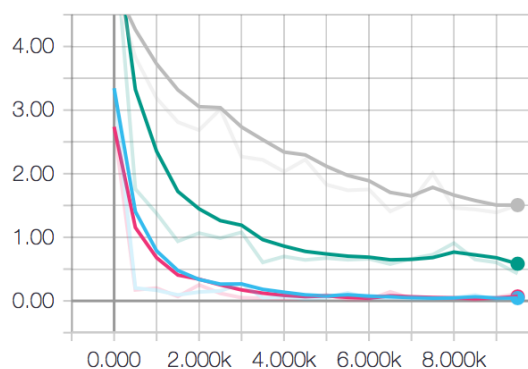
Optimizer: Gradient Descend (Learning Rate = 0.5, 1, 0.01, 0.001)
 Activation Function: Tanh
 Hidden Layer 2: 50 Neurons
 Hidden Layer 2: 50 Neurons
 Mini-batch size: 64

نتایج به صورت زیر است:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss



Name	Smoothed	Value	Step	Time	Relative
Monitoring4	0.9786	1.000	4.500k	Wed Nov 14, 21:19:44	4s
mnist/Monitoring6_LR0.001	0.3371	0.3281	4.500k	Wed Nov 14, 21:43:34	4s
mnist/Monitoring6_LR0.01	0.7679	0.7969	4.500k	Wed Nov 14, 21:43:00	4s
mnist/Monitoring6_LR1	0.9891	1.000	4.500k	Wed Nov 14, 21:39:21	4s

برای ۱۰۰۰۰ حرکت، صحت نیز به صورت زیر است:

Test Accuracy for Learning Rate = 1 : 0.96
 Test Accuracy for Learning Rate = 0.5 : 0.96
 Test Accuracy for Learning Rate = 0.01 : 0.91
 Test Accuracy for Learning Rate = 0.001 : 0.68

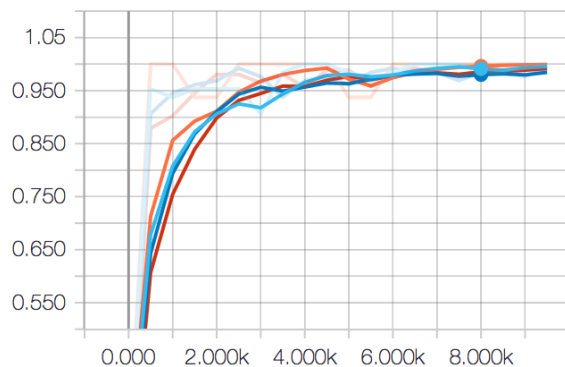
که نتیجه می‌گیریم احتمالاً بخاطر عدم وجود Local Minimum ها و Saddle Point های زیاد و سادگی شبکه، هرچه Learning Rate بیشتر باشد سرعت همگرایی و نتیجه یادگیری شبکه بهتر است.

سوال هفتم:

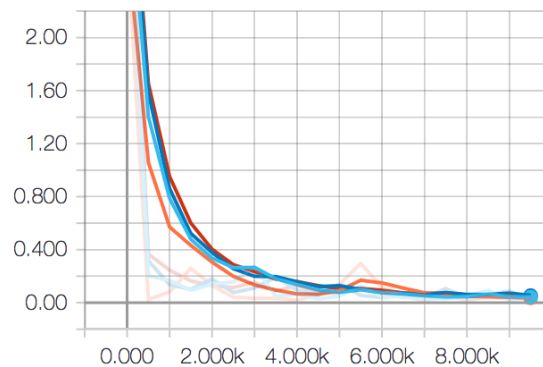
Optimizer: Gradient Descend (Learning Rate = 0.5)
 Activation Function: Tanh
 Hidden Layer 2: 50 Neurons
 Hidden Layer 2: 50 Neurons
 Mini-batch size: 64, 16, 128, 256

نتایج به صورت زیر است:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss



	Name	Smoothed	Value	Step	Time	Relative
●	Monitoring4	0.9910	0.9844	8.000k	Wed Nov 14, 21:19:48	8s
●	Monitoring7_batch128	0.9799	0.9844	8.000k	Wed Nov 14, 21:54:08	10s
●	Monitoring7_batch16	0.9962	1.000	8.000k	Wed Nov 14, 21:52:41	5s
●	Monitoring7_batch256	0.9852	0.9922	8.000k	Wed Nov 14, 21:54:48	16s

برای ۱۰۰۰۰ حرکت، صحت نیز به صورت زیر است:

Test Accuracy for Mini-batch Size = 16 : 0.92
 Test Accuracy for Mini-batch Size = 64 : 0.96
 Test Accuracy for Mini-batch Size = 128 : 0.91
 Test Accuracy for Mini-batch Size = 256 : 0.97

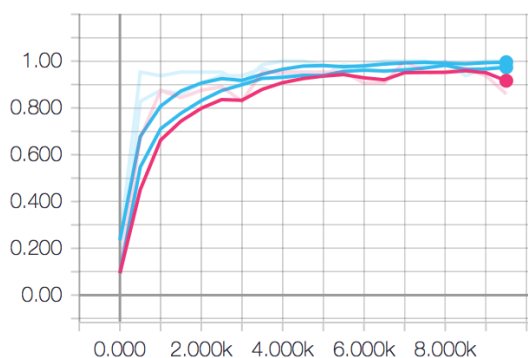
که ملاحظه می شود سایز Mini-batch تاثیر چندانی در آموزش این شبکه ساده ندارد، اما با افزایش سایز آموزش شبکه وقت گیرتر می شود.

سوال هشتم:

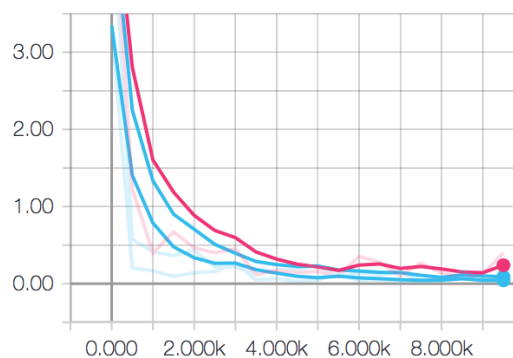
Optimizer: RMSprop (Learning Rate = 0.001), Adam (Learning Rate = 0.001)
 Activation Function: Tanh
 Hidden Layer 2: 50 Neurons
 Hidden Layer 2: 50 Neurons
 Mini-batch size: 64

نتایج به صورت زیر است:

Accuracy
tag: Save/Accuracy



Loss
tag: Save/Loss



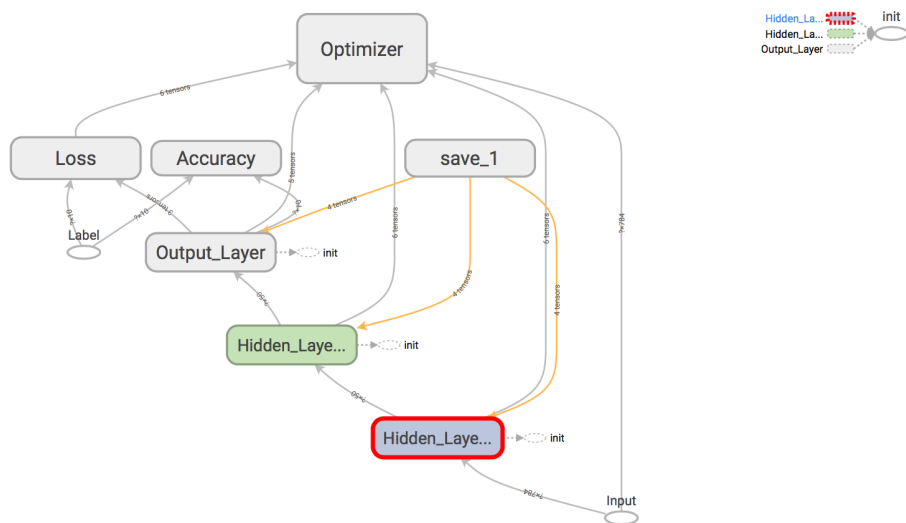
Name	Smoothed	Value	Step	Time	Relative
Monitoring4	0.9956	1.000	9.500k	Wed Nov 14, 21:19:49	9s
Monitoring8_Adam	0.9159	0.8594	9.500k	Wed Nov 14, 22:03:55	11s
Monitoring8_RMSProp	0.9732	0.9844	9.500k	Wed Nov 14, 22:01:46	10s

(نمودار بالایی در Accuracy و نمودار پایینی در loss مربوط به GD است)
 صحت نیز به صورت زیر است:

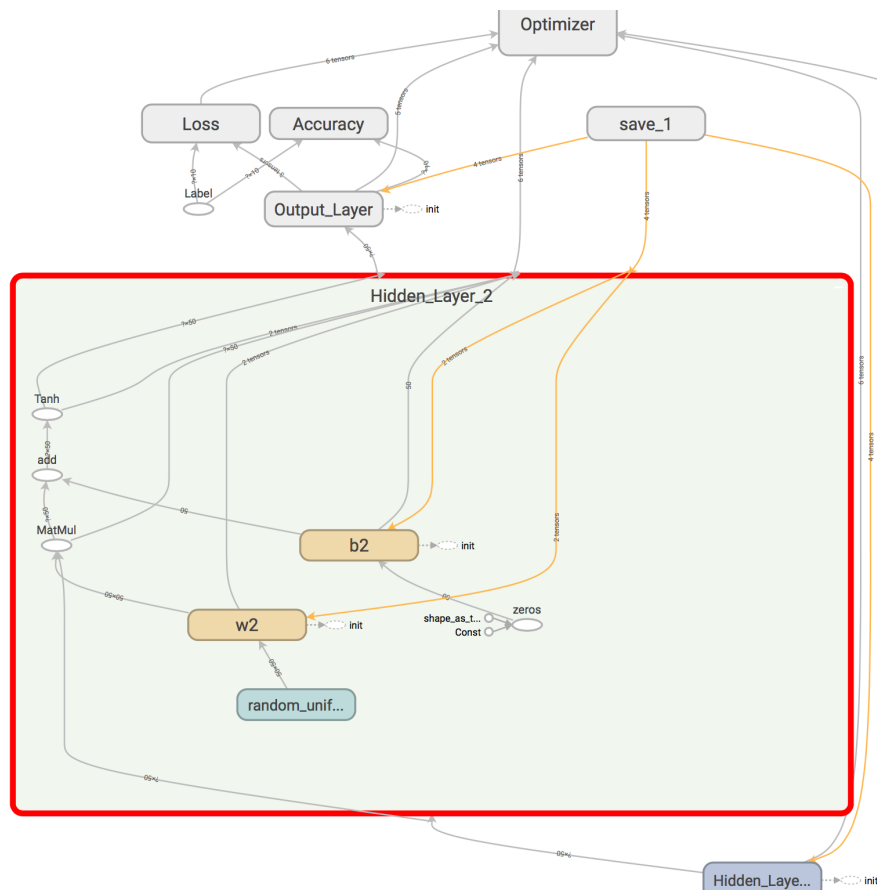
Test Accuracy for Gradient Descend : 0.96
 Test Accuracy for RMSProp : 0.96
 Test Accuracy for Adam : 0.92

که ملاحظه می شود در اینجا (احتمالا به دلیل سادگی شبکه) Gradient Descend عملکرد بهتری دارد.

سوال نهم:



و برای مثال یک لایه مخفی بدین شکل است:



پایان گزارش
