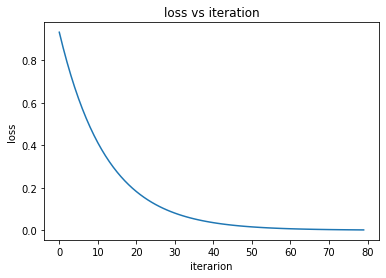
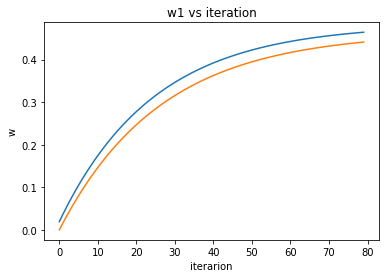
**Toy nets, no bias, initial weights with normal distribution:**

**"Toy net" of one fully connected layer, one value input**

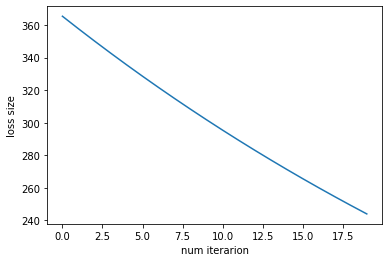
Training process:

Display the exponential rate of the weight convergence.



**"Toy net" of one fully connected layer, input**

Training process:



**"Toy net" of tow fully connected layers**

**Dimensions**:

**First layer:**

**Second layer:**

Training process:

תמונה שמכילה ריבוע

התיאור נוצר באופן אוטומטי

The initial weights of each layer were saved before the training.

Using the initial weights of the first layers and the final of the second the loss is: 1.096

Using the initial weights of the second layers and the final of the first the loss is: 0.001

As it seems, the first layer is more significant for the net.

Freeze every time one layer and train the net:

The loss as function of the iteration for freezing the first layer:

תמונה שמכילה ריבוע

התיאור נוצר באופן אוטומטי

The loss as function of the iteration for freezing the second layer:תמונה שמכילה ריבוע

התיאור נוצר באופן אוטומטי

As it seems, it's possible to train the net (and to achieve ideal results) using only one layer.

The contribution of each layer for different dimensions to the layers.

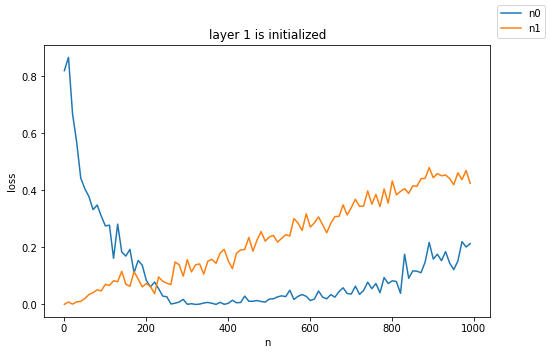
While the dimensions of layer 0 are and the dimensions of layer 1 are: :

The contribution of the first layer:

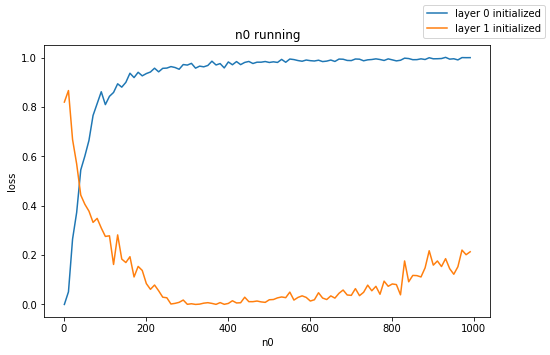
**תמונה שמכילה טקסט, צילום מסך, מסגרת תמונה

התיאור נוצר באופן אוטומטי**

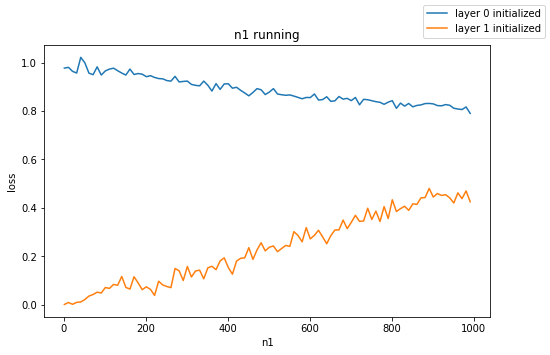
The contribution of the second layer:



Comparison between the layers while for different values for



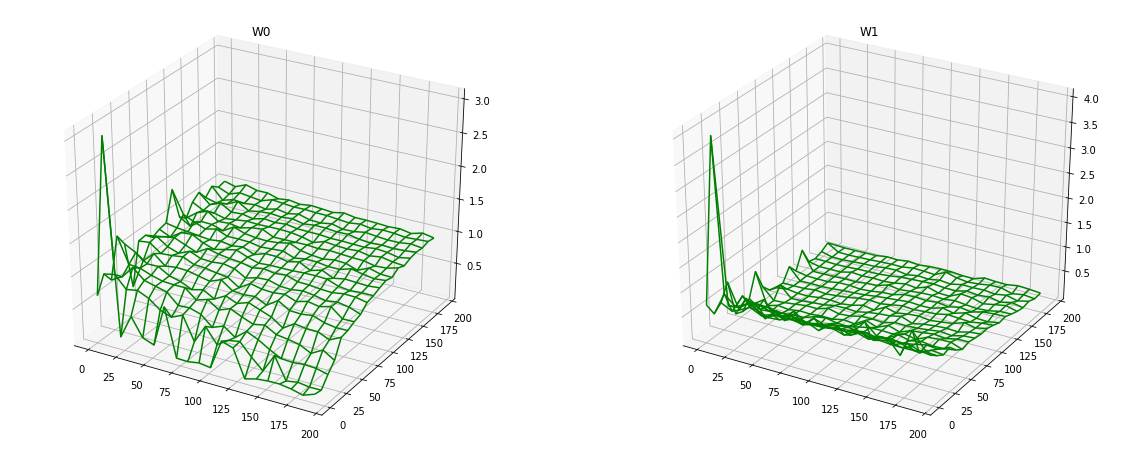
Comparison between the layers while for different values for



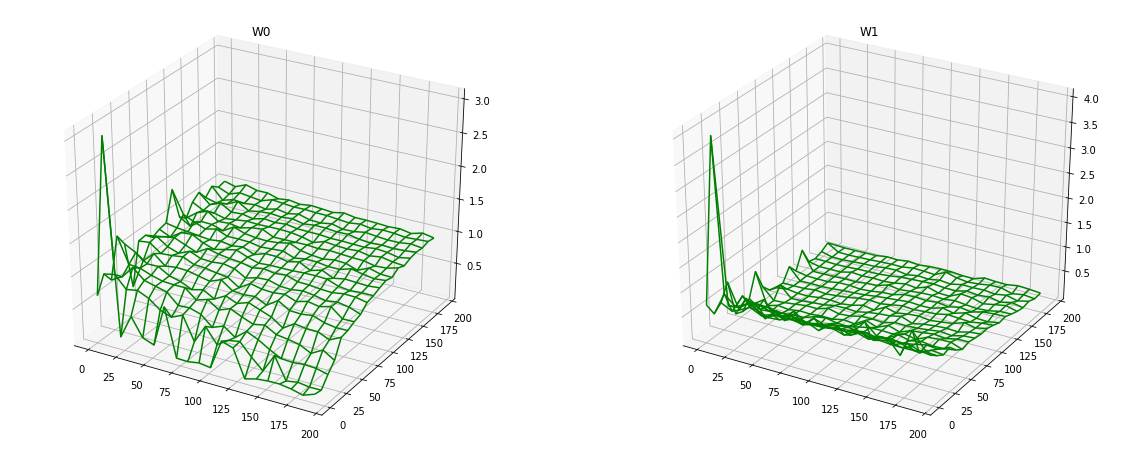
As it seems, the ratio between the size of the layers affects about the ratio between the contribution of them.

**The same net for**

The contribution of the first layer:



The contribution of the second layer:



Comparison between SGD and Adam optimizer of the contribution of each layer for same dimensions (n) for the two layers:

LogLog scaling:

תמונה שמכילה חלון, שוג'י, מקורה

התיאור נוצר באופן אוטומטי

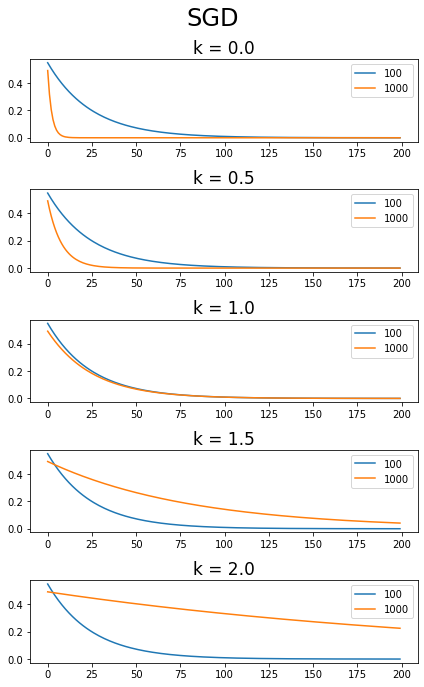
Log10 scaling:

תמונה שמכילה טקסט, שוג'י

התיאור נוצר באופן אוטומטי

Searching for the ideal k parameter (while learning rate is: ) for to insure the same convergence rate to different dimensions of net:

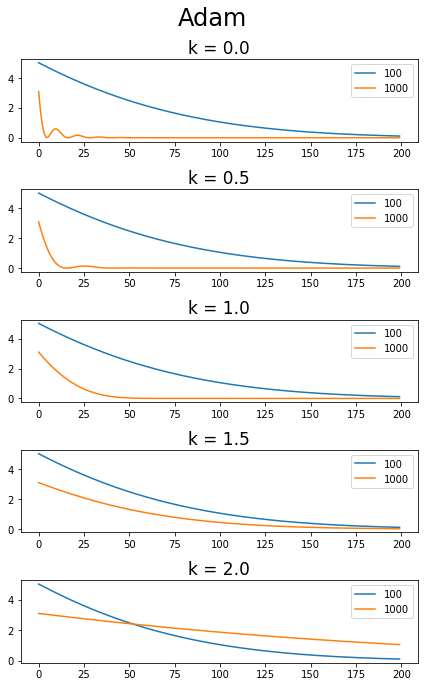
k = 1.0 for SGD optimizer:



תמונה שמכילה שוג'י

התיאור נוצר באופן אוטומטי

k = 1.5 for Adam optimizer:

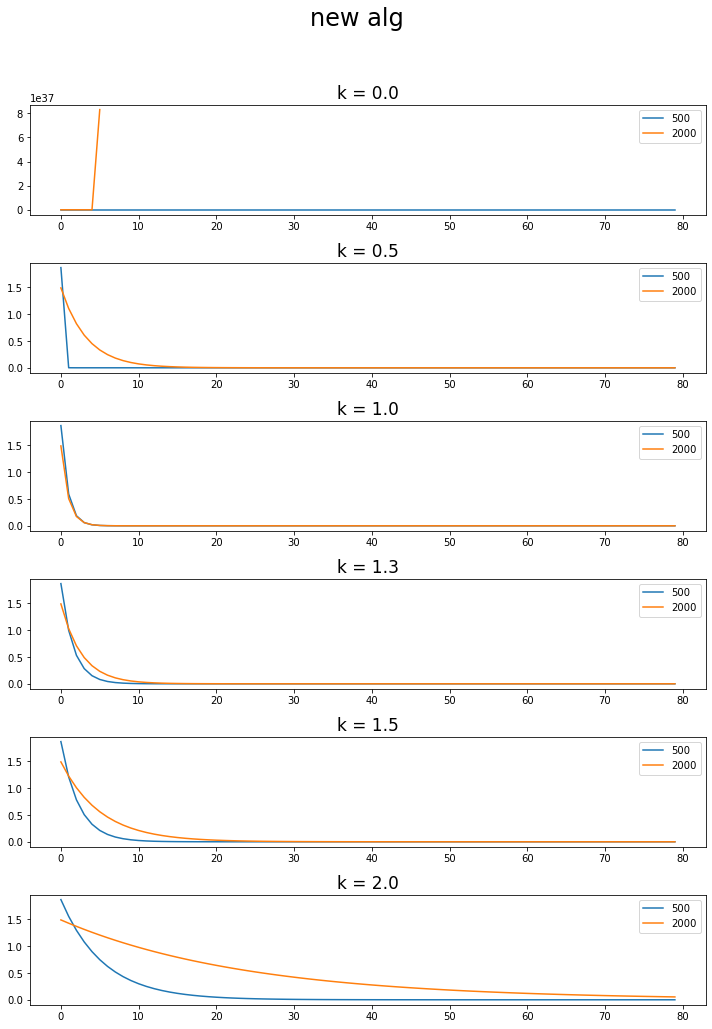


תמונה שמכילה טקסט, שוג'י

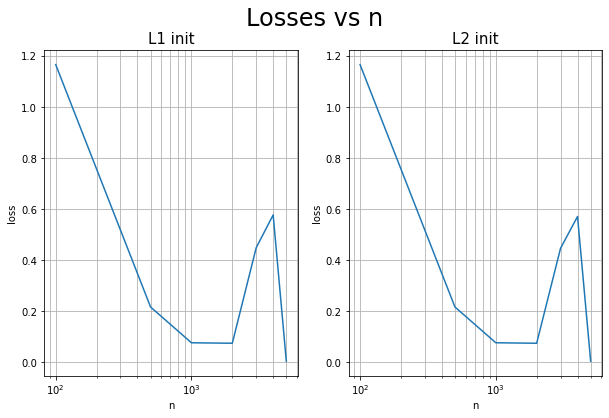
התיאור נוצר באופן אוטומטי

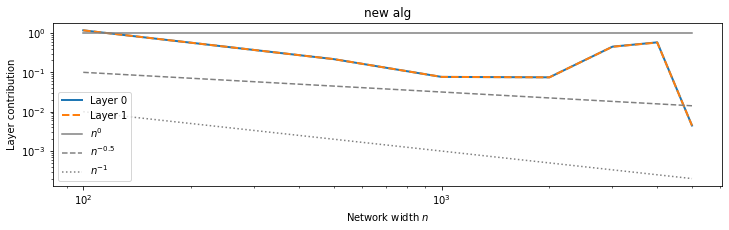
**Using "new\_alg"**

Finding the ideal k:



k=1.0

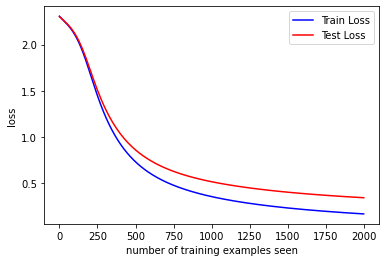
Comparison the contribution of each layer using new alg: 



**MNIST**

**Using SGD:**

Training:



the contribution of every layer:

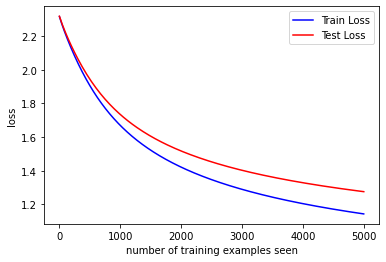
fc1 contribution is: tensor(1.5483)

fc2 contribution is: tensor(1.0275)

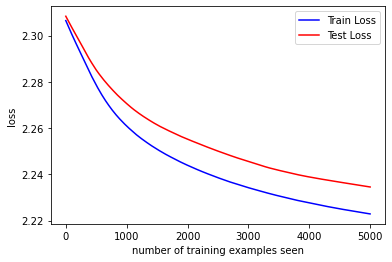
conv1 contribution is: tensor(0.4434)

conv2 contribution is: tensor(1.2716)

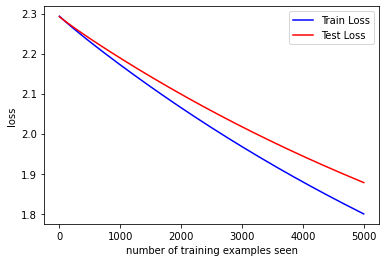
Training only fc1 layer:



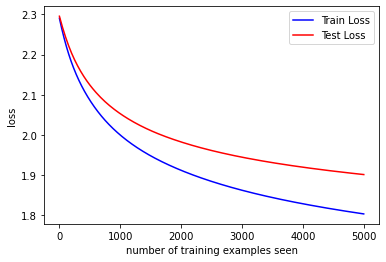
**Training only conv1 layer:**



Training only fc2 layer:



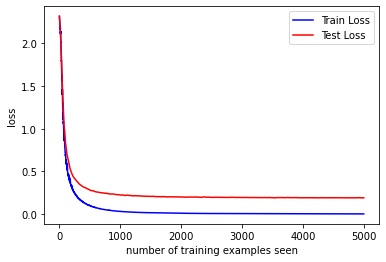
Training only conv2 layer:



As it seems, not all the four layer have the same contribution and it isn't possible to achieve ideal loss using only one layer for every layer.

**MNIST with new\_alg**

Train the net:



The difference between the contribution of the layers:

fc1 contribution is: tensor(1.1091)

fc2 contribution is: tensor(1.0625)

conv1 contribution is: tensor(5.2736)

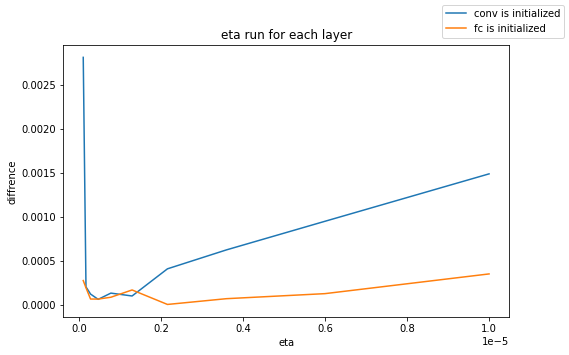
conv2 contribution is: tensor(2.9156)

TODO: not achieved the goal

new\_alg, comparison eta parameter (learning rate) for equal contribution between the layers for one iteration.

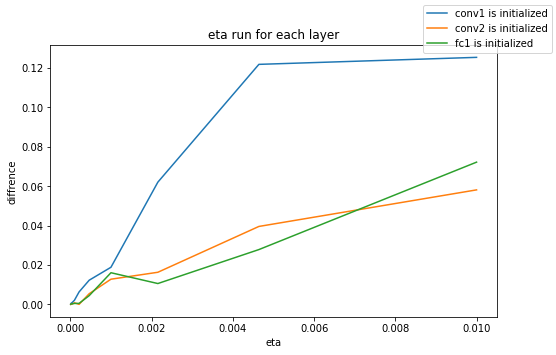
Net: one conv2d layer, one fc layer

The difference between the contribution to



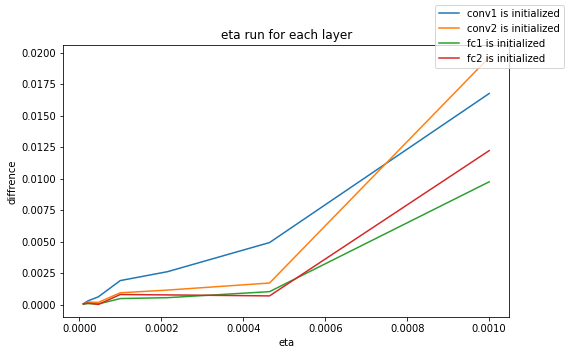
Net: two conv2d layers, one fc layer

The difference between the contribution to :



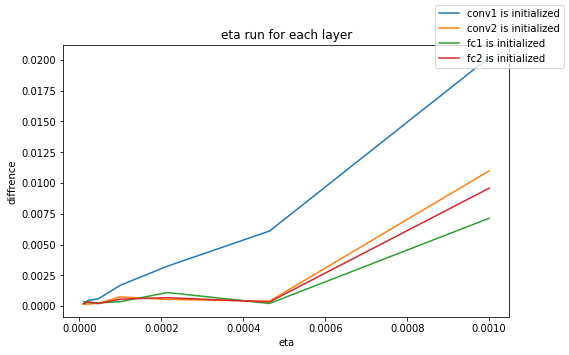
Net: two conv2d layers, two fc layers

The difference between the contribution to :



Net: two conv2d layers, two fc layers with an activation function (tanh ) in the last layer

The difference between the contribution to :



As it seems the equal contribution was achieved for .