Homework 7 : Weight Initialization and Batch Normalization

Dead line:

TODO

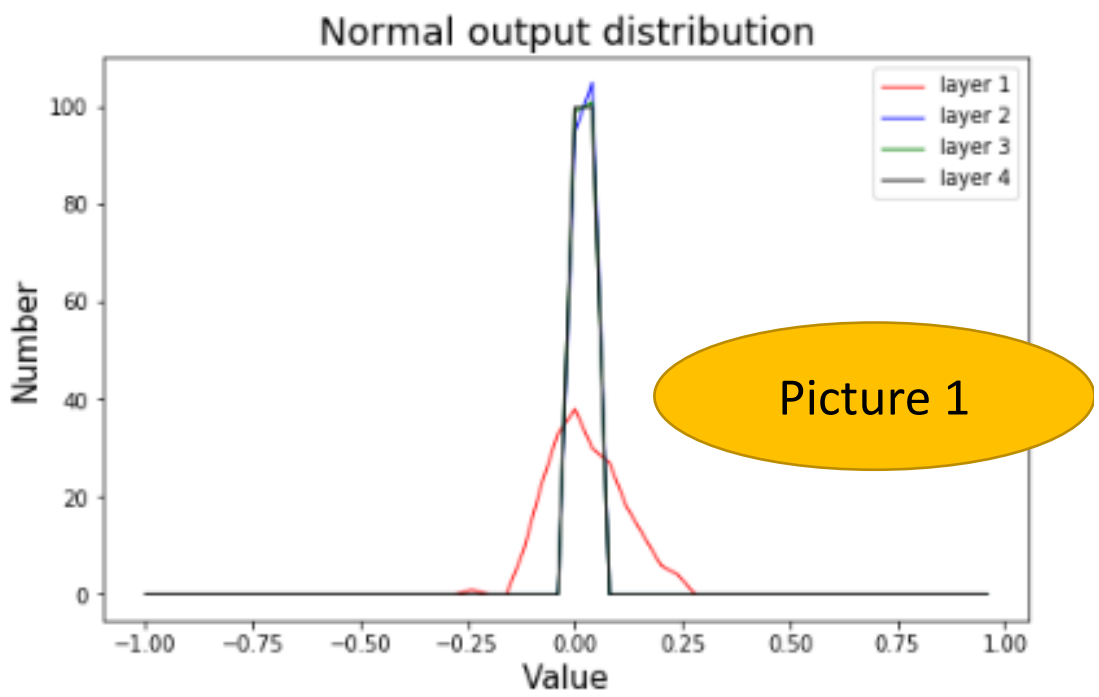
- 1. Plot the value distribution picture of output and output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)
- 2. Plot the variance value picture of output and output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)
- 3. Plot the variance value picture of output and output gradient in very layer. (without and with batchNorm in every layer)
- 4. Plot the variance value picture of output in very layer. (In two different weight initialization method, which are normal and Xavier weight initialization. Every layer has batchNorm layer)

1. Plot the value distribution picture of output and output gradient in very layer. (In three different weight initialization method, which are normal, Xavier and orthogonal weight initialization)

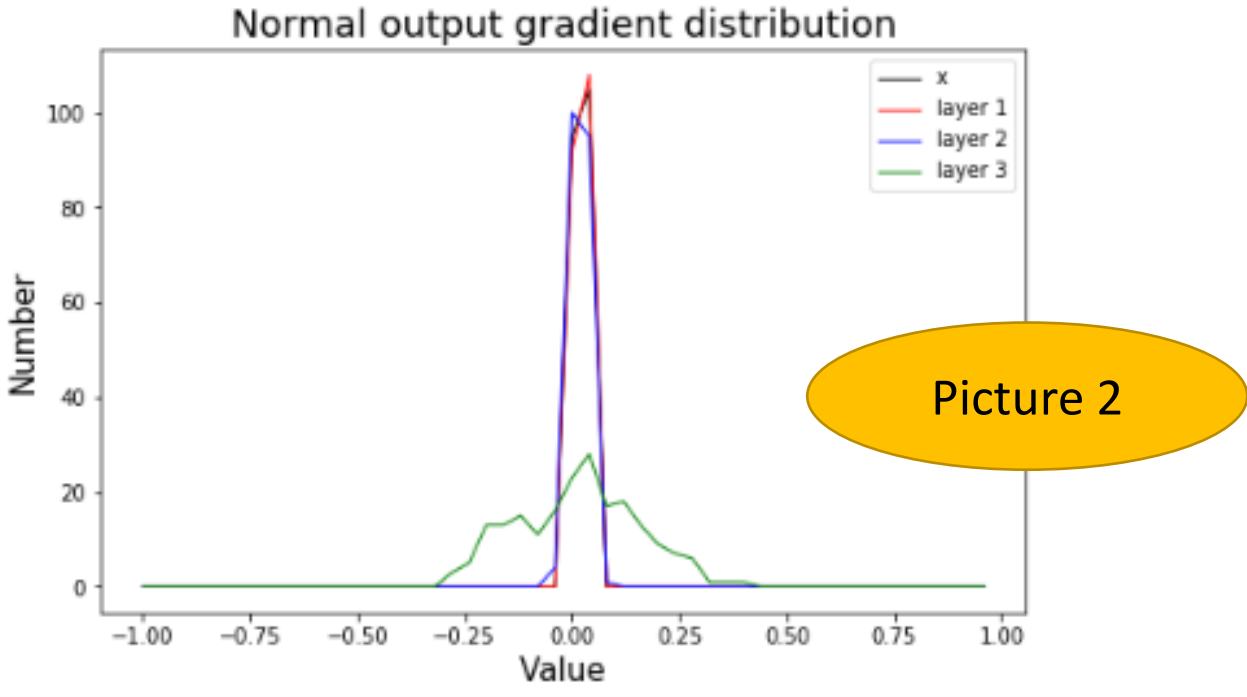
- You may get a four layers model which activation function is tanh
- Use the given input x and forward the model.
- Calculate the variance of output and variance of output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)
- Plot the value distribution picture of output and gradient of output in very layer (In two different weight initialization methods, which are normal and Xavier weight initialization)

1. Plot the value distribution picture of output and output gradient in very layer. (In two different weight initialization method, which are normal and Xavier weight initialization)

- In normal weight initialization



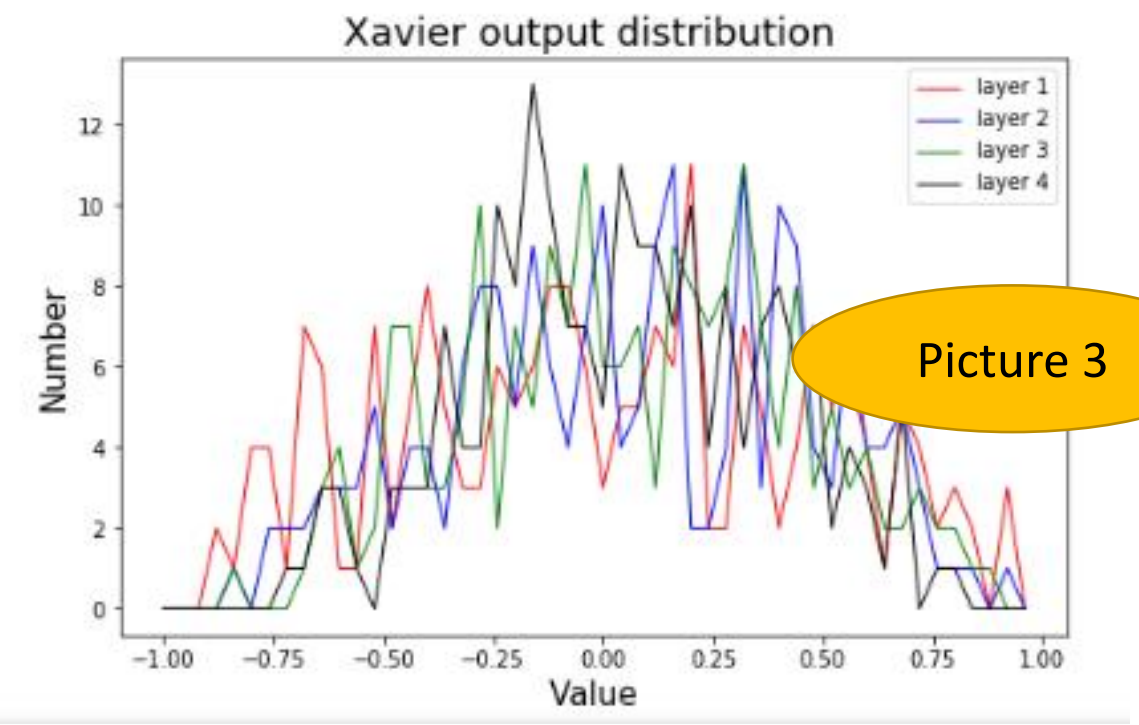
Picture 1



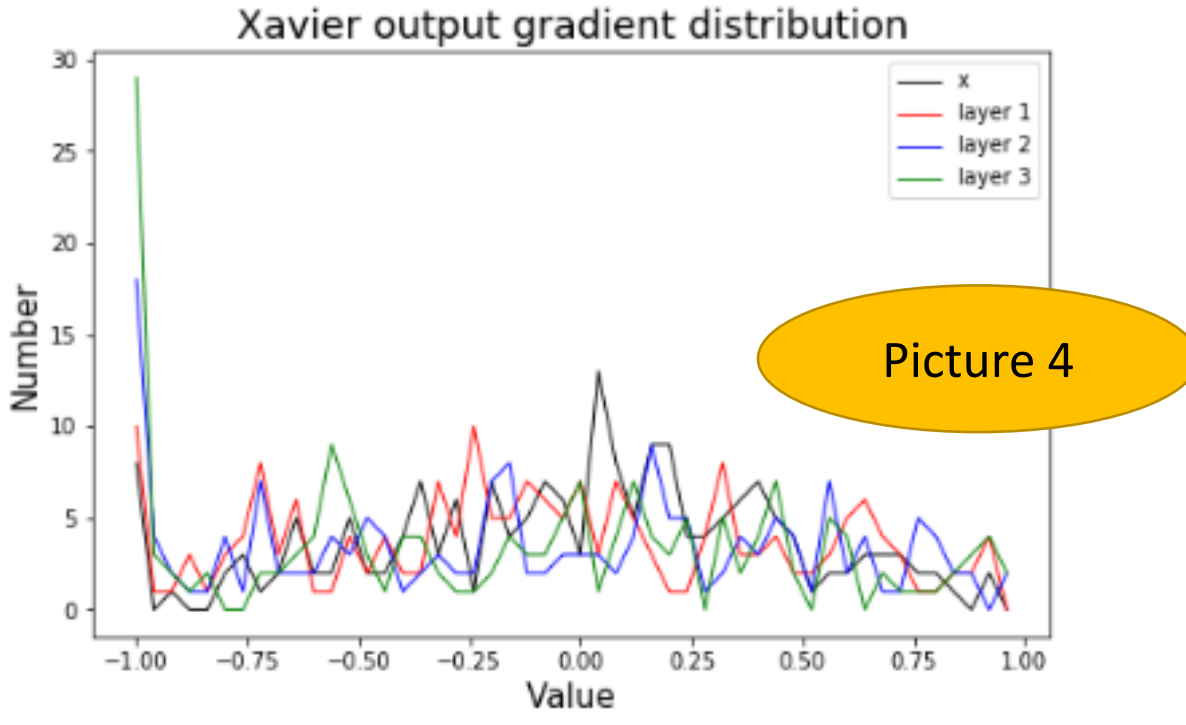
Picture 2

1. Plot the value distribution picture of output and output gradient in very layer. (In three different weight initialization method, which are normal, Xavier and orthogonal weight initialization)

- In Xavier weight initialization



Picture 3



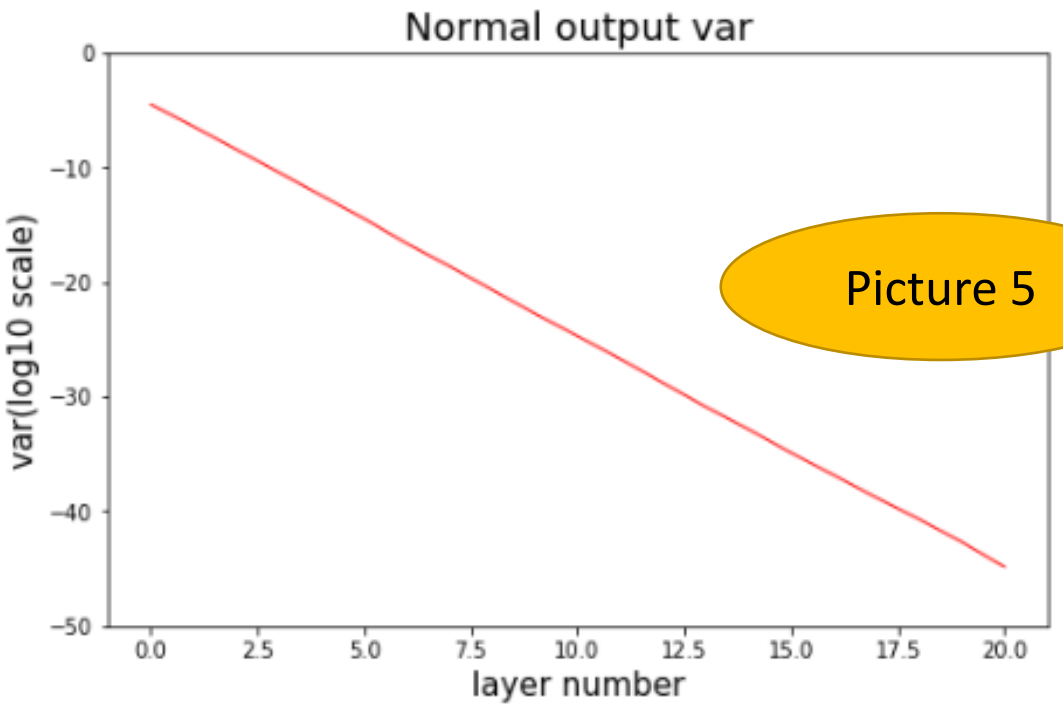
Picture 4

2. Plot the variance value picture of output and output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)

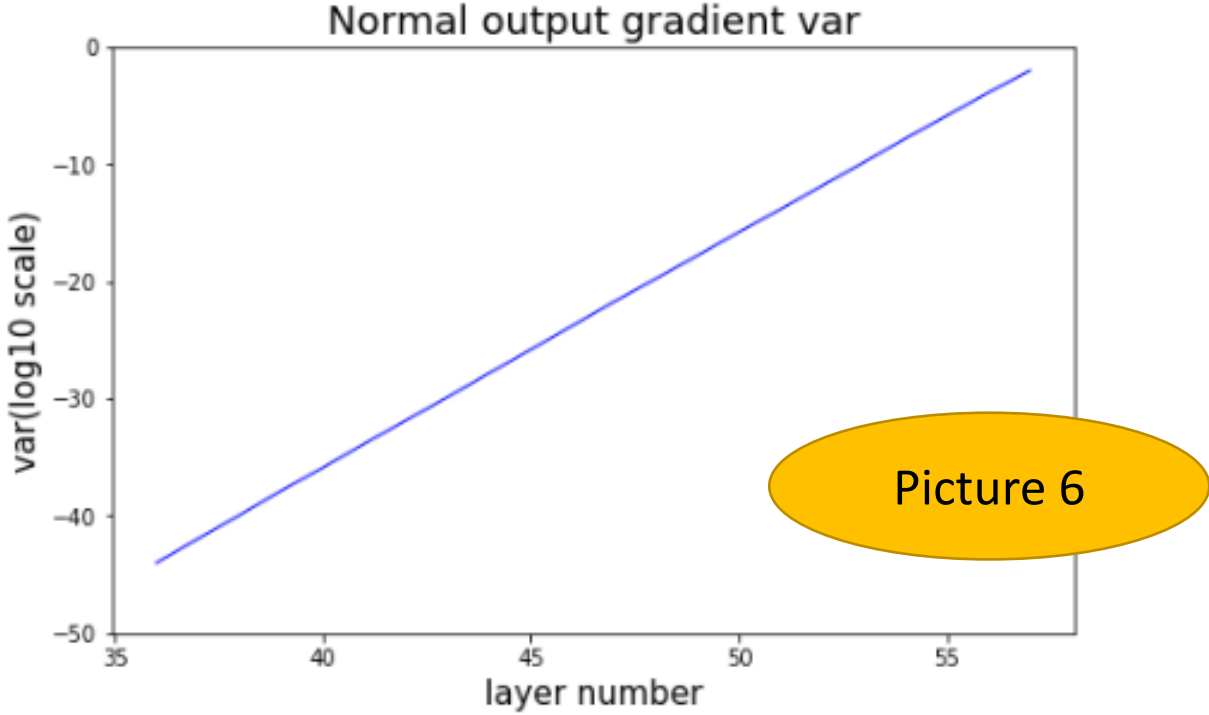
- You may get a 60 layers model which activation function is tanh
- Use the given input x and forward the model.
- Plot the variance value picture of output and output gradient in very layer (In two different weight initialization methods, which are normal and Xavier weight initialization)

2. Plot the variance value picture of output and output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)

- In normal weight initialization



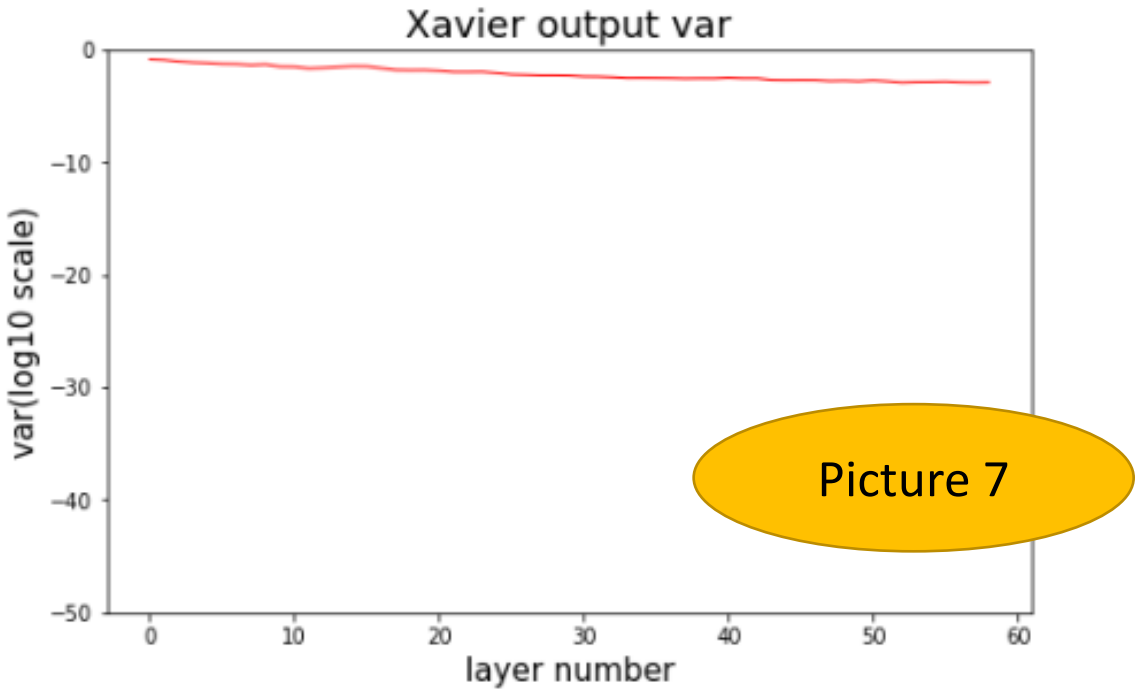
Picture 5



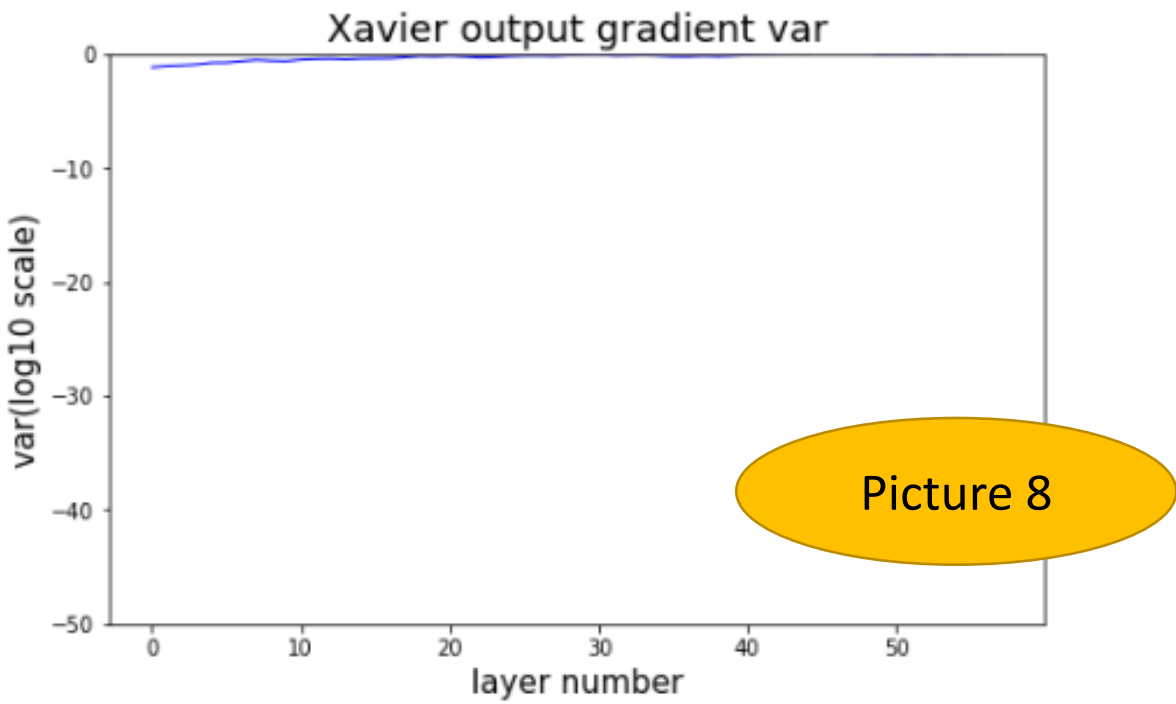
Picture 6

2. Plot the variance value picture of output and output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)

- In Xavier weight initialization



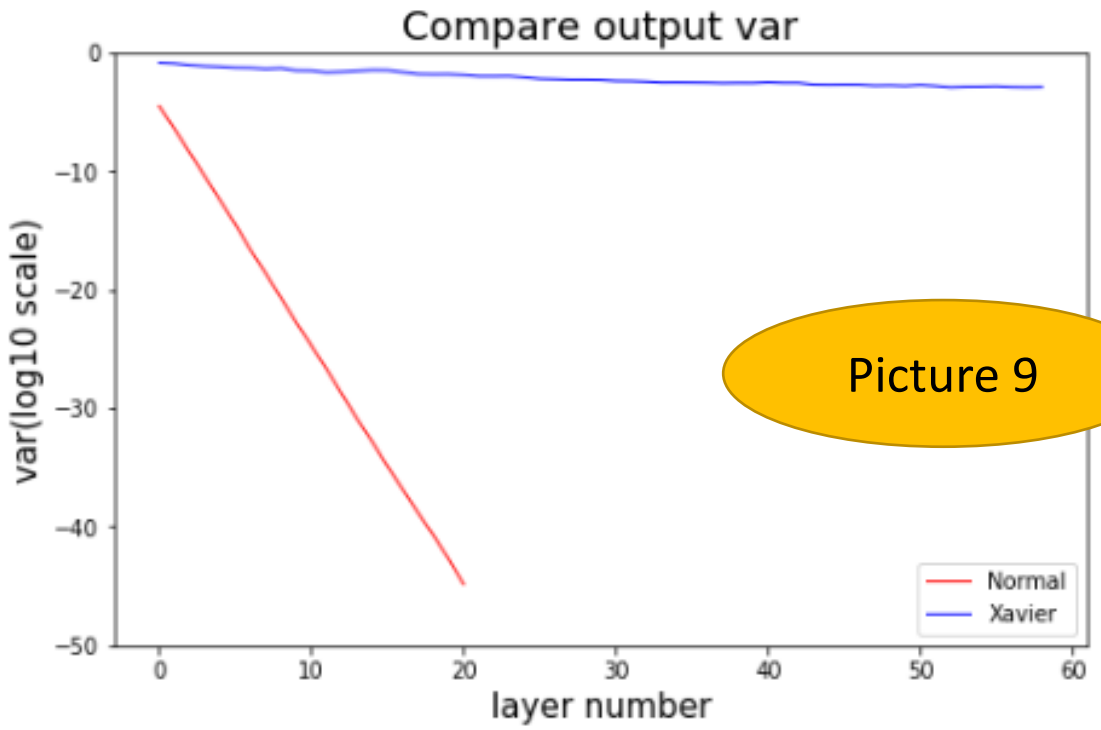
Picture 7



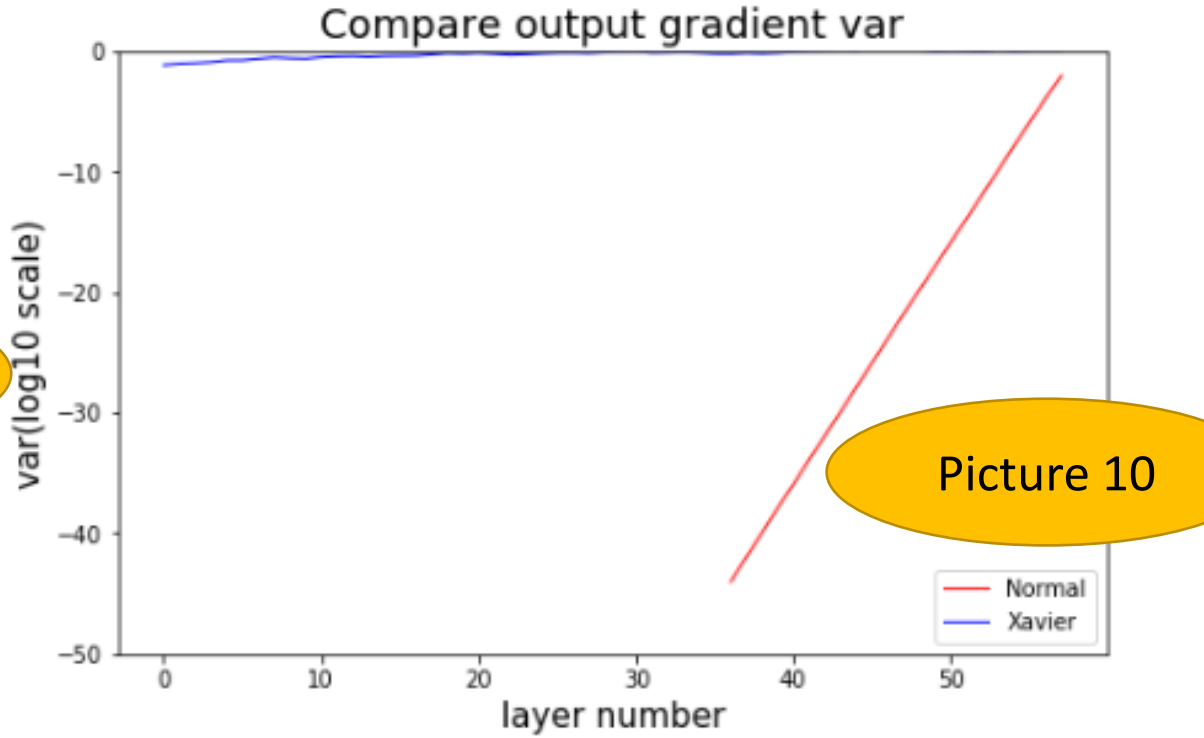
Picture 8

2. Plot the variance value picture of output and output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)

- Compare two results



Picture 9



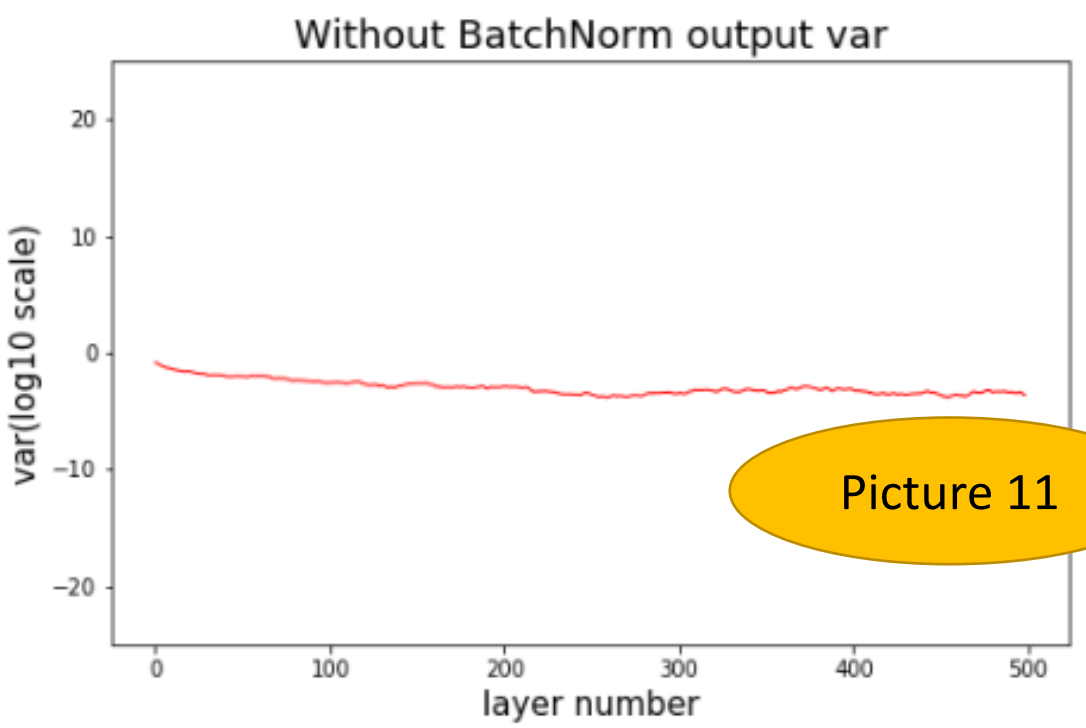
Picture 10

3. Plot the variance value picture of output and output gradient in very layer. (without and with batchNorm in every layer)

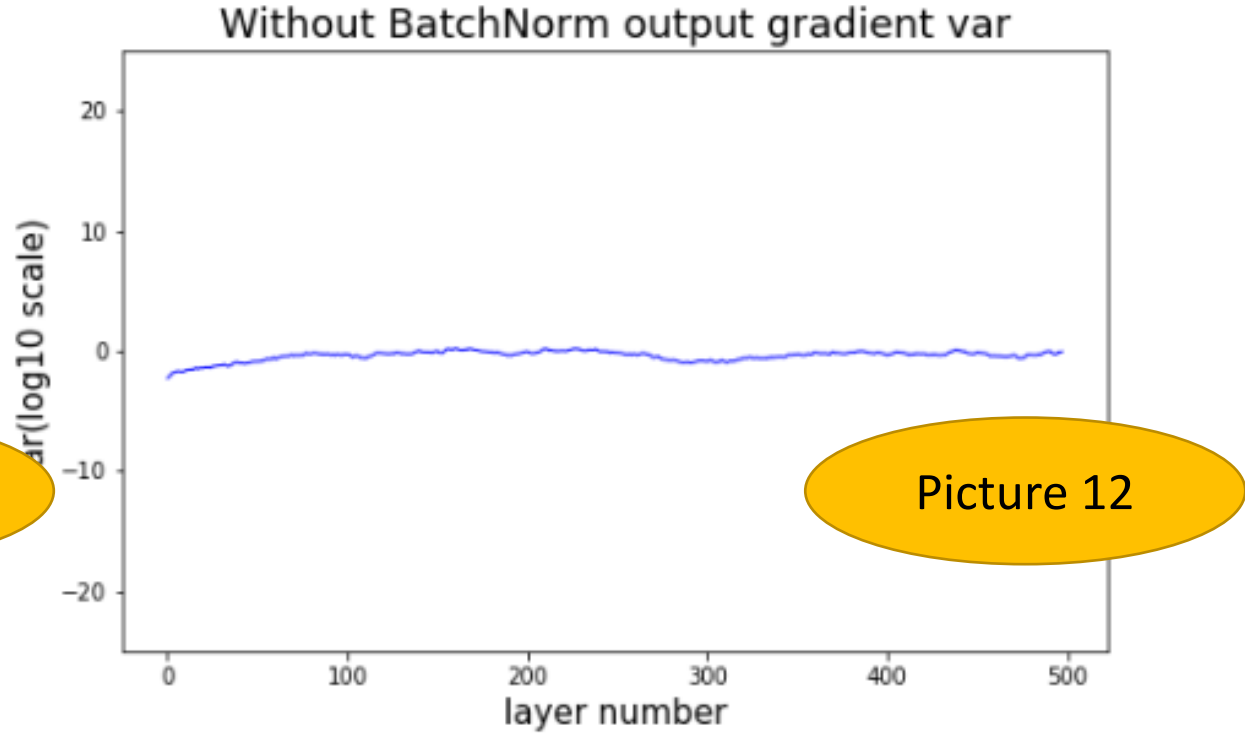
- You may get a 500 layers model which activation function is tanh
- Use the given input x and forward the model.
- Plot the variance value picture of output and output gradient in very layer (without and with batchNorm in every layer)

3. Plot the variance value picture of output and output gradient in very layer. (without and with batchNorm in every layer)

- Without BatchNorm layer



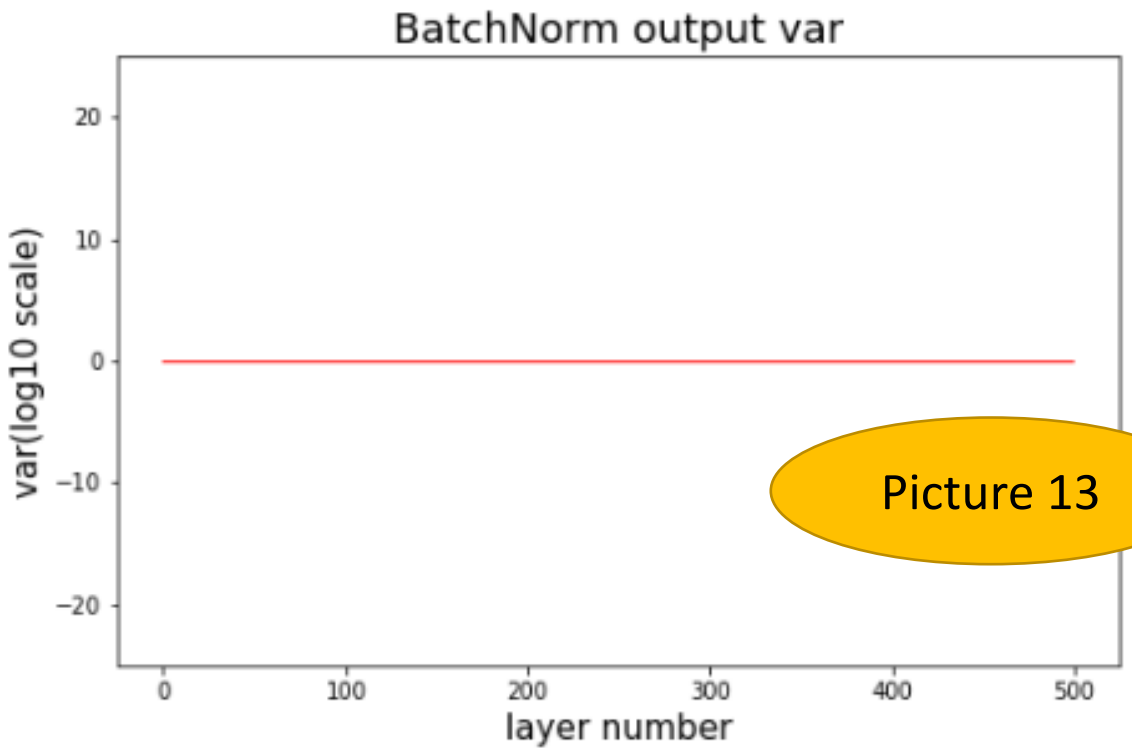
Picture 11



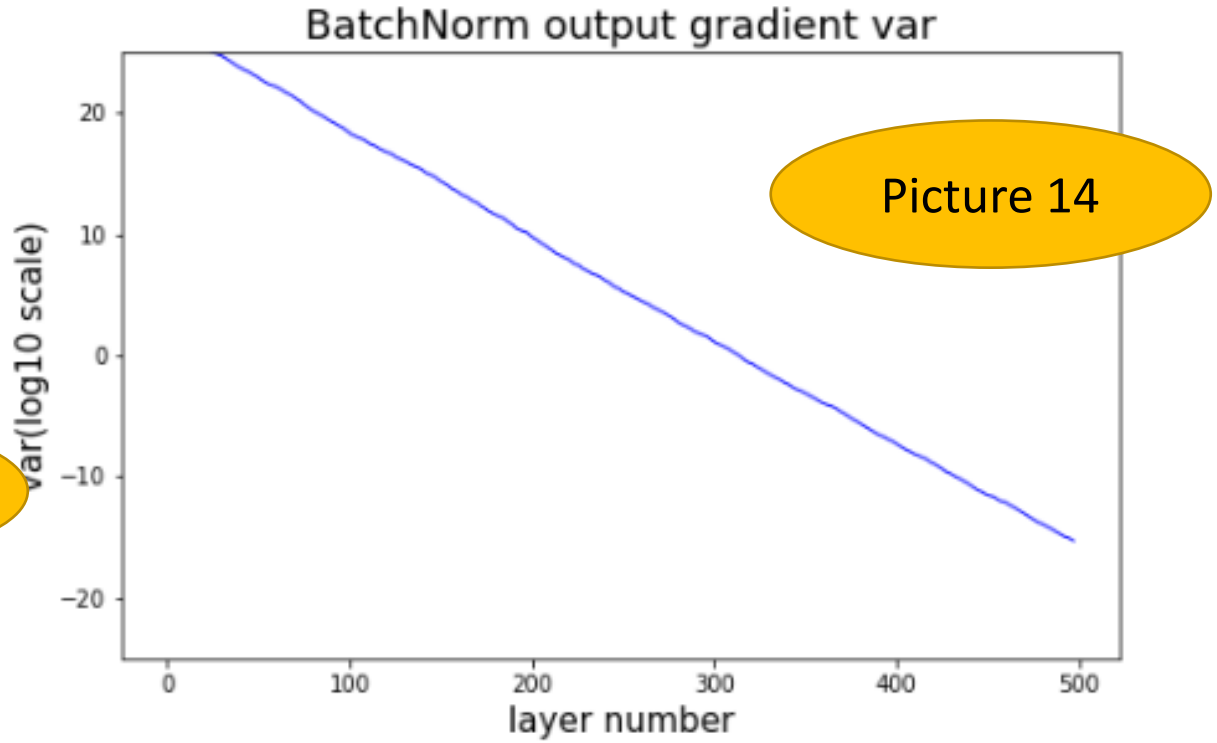
Picture 12

3. Plot the variance value picture of output and output gradient in very layer. (without and with batchNorm in every layer)

- With BatchNorm layer



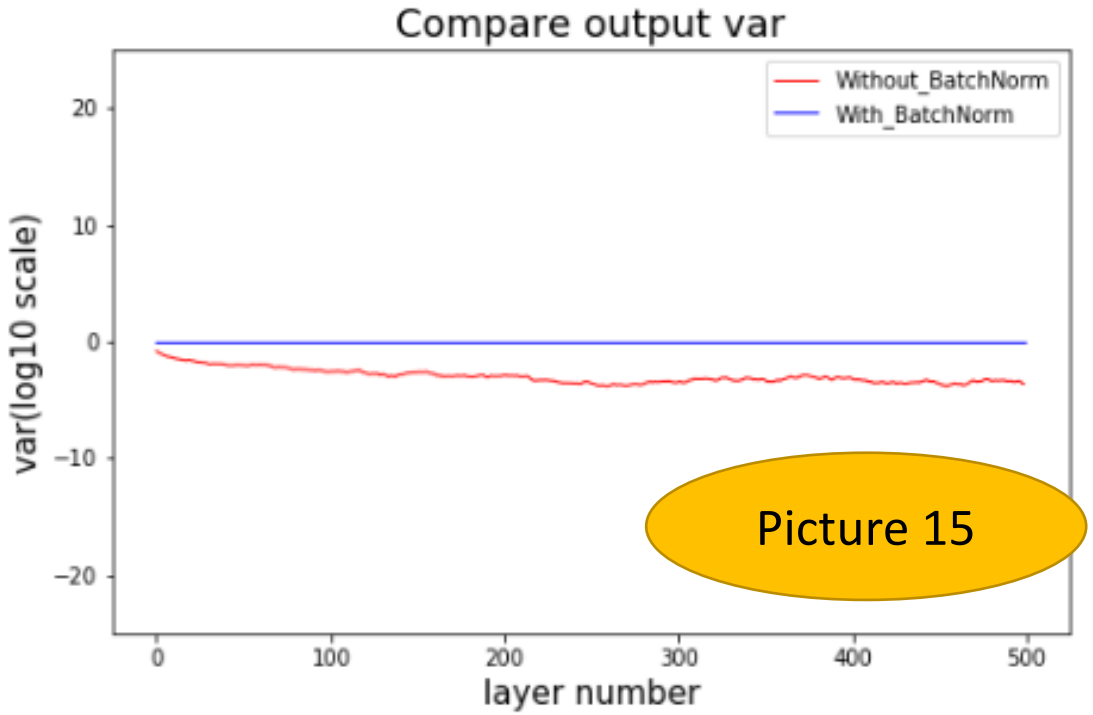
Picture 13



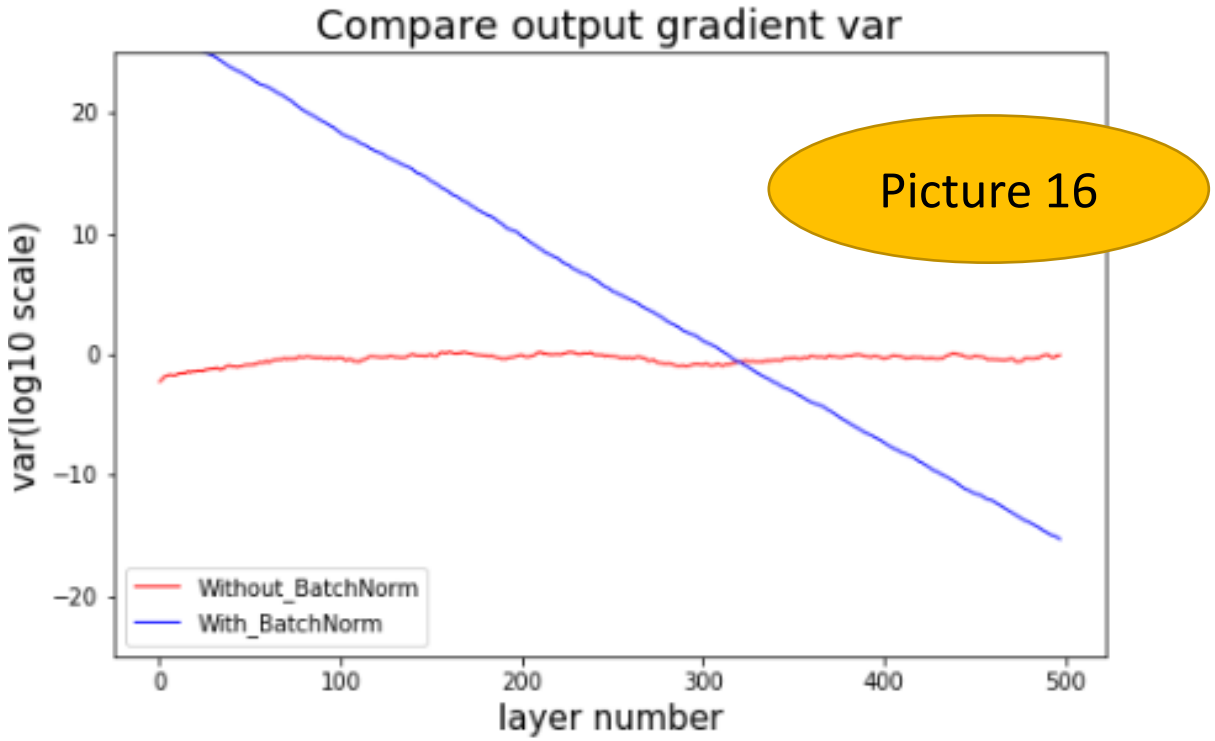
Picture 14

3. Plot the variance value picture of output and output gradient in very layer. (without and with batchNorm in every layer)

- Compare two results



Picture 15



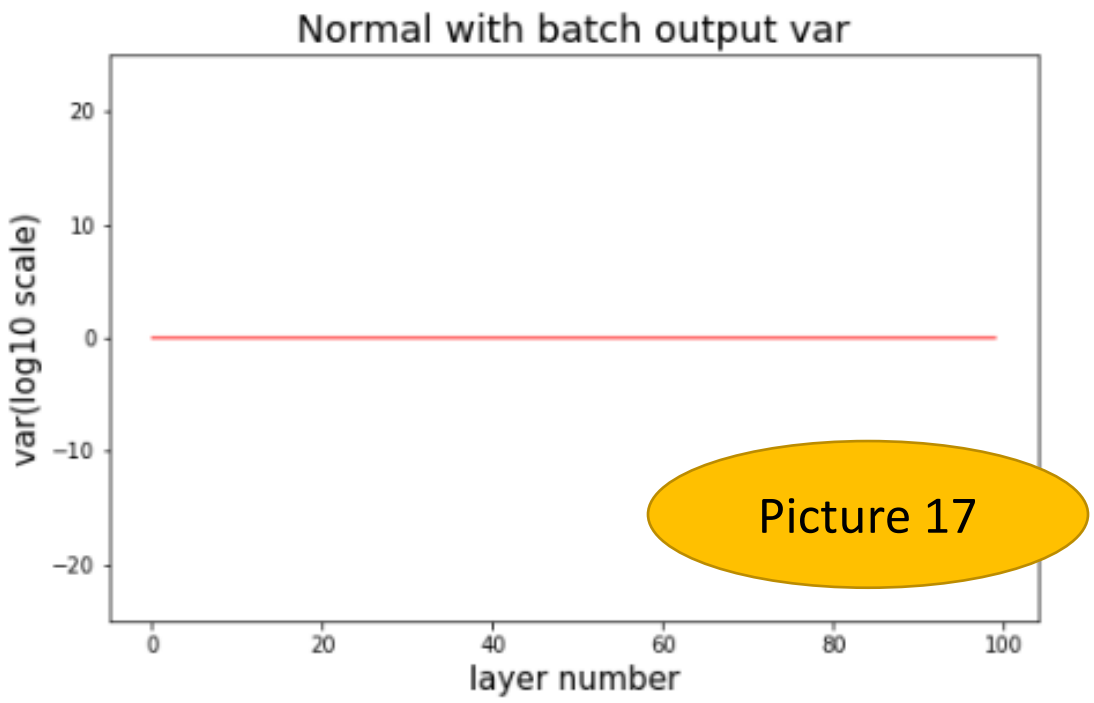
Picture 16

4. Plot the variance value picture of output in very layer. (In two different weight initialization method, which are normal and Xavier weight initialization. Every layer has batchNorm layer)

- You may get a 100 layers model which activation function is tanh and each layer has batchnorm
- Use the given input x and forward the model.
- Plot the variance value picture of output in very layer (In two different weight initialization method, which are normal and Xavier weight initialization. Every layer has batchNorm layer)

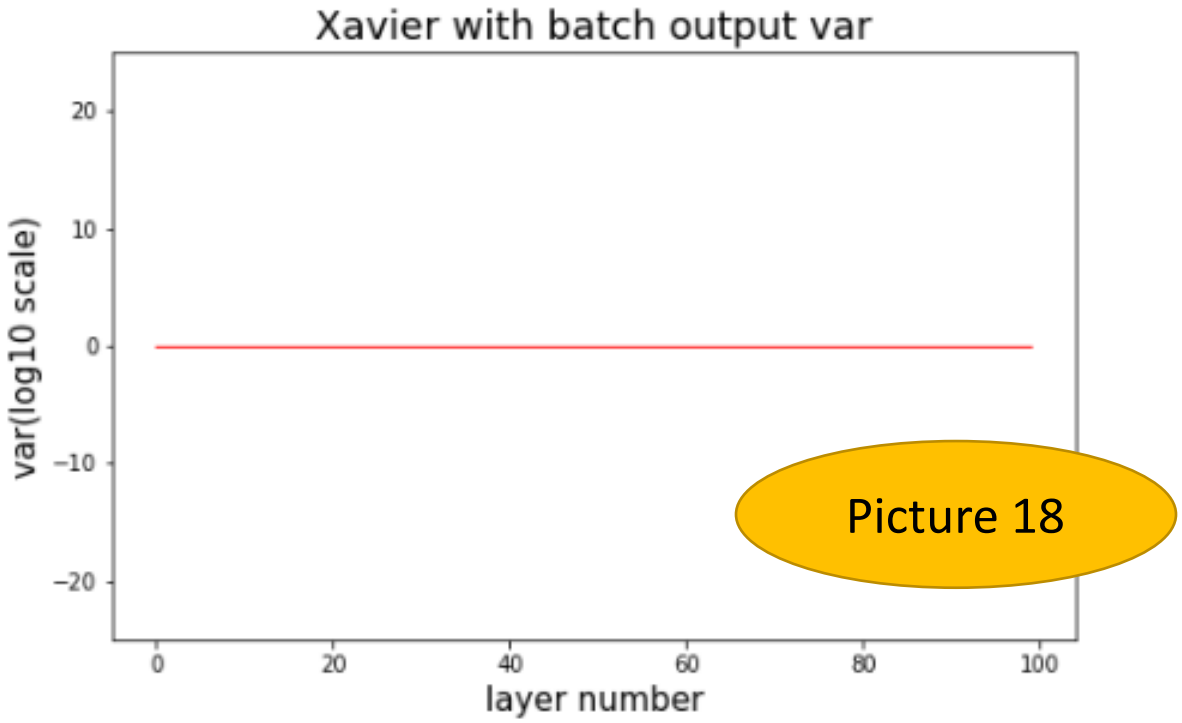
4. Plot the variance value picture of output in very layer. (In two different weight initialization method, which are normal and Xavier weight initialization. Every layer has batchNorm layer)

- In normal weight initialization



Picture 17

- In Xavier weight initialization



Picture 18

Reference

- Understanding the difficulty of training deep feedforward neural networkshttp://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf(Xavier weight initialization)