1. I have implemented the DNN using SGD. This algorithm works for any number of hidden layers (Even with zero!) . Just edit the hyper parameters as you wish from the **main** function. If you want tune the hyper parameters , give the as lists in the **main** function , and change the key word argument of **dnn** class to **True**. If the tune parameter is set to false, the network chooses a random hyperparameter set.
2. At 3 Hidden layers and 64 hidden units per each hidden layer, check\_grad function returns the following output:

**Check grad: 8.48675979004169e-07**