Machine Learning CSE 4020

Assignment 2

Date: August 30, 2019

Question 1: Compare the performance of two neural networks when one is using a linear activation function while another neural network is using all nonlinear activation function.

Question 2: Using back propagation tries to approximate following function in range [-2π, 2π] and plot the true function and approximated function.

- sin(x)/x
- $\frac{y}{x} = \tan(\frac{x^2+y^2}{a^2})$
- sin(sin(x)+cos(y)) = cos(sin(xy) + cos(x))
- $$\begin{aligned} x(t) &= R\left[(1-k)\cos t + lk\cos\frac{1-k}{k}t\right], \\ y(t) &= R\left[(1-k)\sin t - lk\sin\frac{1-k}{k}t\right]. \end{aligned}$$

Question 3: Perform an experiment to compare between MSE, Cross Entropy, and Hinge Loss. Use MNIST, CIFAR10 or LFW dataset.

Question 4: Propose a method to use the Neural Network for time series data.

Question 5: Compare the effect of initialization of weight of neural network with following distributions:

- Standard Normal Distribution
- Standard t Distribution
- Uniform Distribution
- Exponential Distribution
- Laplace Distribution

Question 6: Derive the update equation in back propagation when cost function includes L2 and L1 regularization terms.

Question 7: In the gradient descend, we use first order partial derivatives of Loss function to update the parameters. List the pros and cons if we use second order partial derivatives.

Question 8: Implement the binary neural network [all outputs are 0 or 1] and perform a classification task for some small dataset and report the inferences.

Question 9: If we have used dropout during training, then the resultant model works like an ensemble of various sparse neural networks in which output of final model is equivalent to the geometric mean of output of all sparse network. Prove it.

Question 10: Discuss the expressiveness of neural network with large no of neurons in few hidden layers vs. less number of neuron in large number of hidden layers.