**Foundations of Deep Learning – Homework Assignment #3**Adi Album & Tomer Epshtein

**Part 3: (4)**

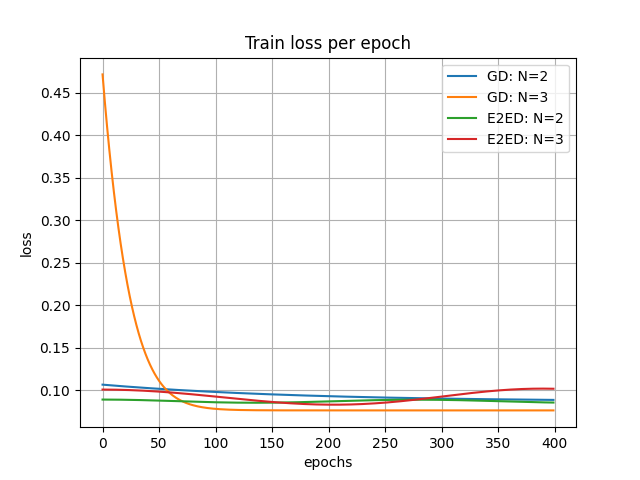
Question:

On a scalar regression dataset of your choice, train a depth linear neural network (with hidden widths no smaller that the minimum between input dimension and output dimension) by minimizing loss via (full batch) gradient descent with small learning rate and initialization close to zero. Compare the trajectory taken by the end-to-end matrix to that obtained by directly applying the (discrete version of the) end-to-end dynamics to a linear model:

Where is the learning rate used for gradient descent over the linear neural network. Repeat the experiment with depths and

Solution:

The training procedure can be viewed in the following plots:



Convergence of all four models: GD with and E2ED (End-to-end dynamics) with is achieved.

Additionally, because our network is a Linear Neural Network with input dimension and output dimension , the end-to-end matrix is a matrix. I.e. a single real value. We can see convergence to global minimum, with different learning rates, in the following graph, where all four models converge to and end-to-end matrix

