Homework 2: Neural Differential Equation (NDE)

Deadline Friday June 21, 17:00

Notebook: Homework_2_pre.ipynb

In this exercise, you will explore the solution method of a neural differential equation, given by

$$\frac{d\mathbf{z}}{dt} = \mathbf{f}(\mathbf{z}, t, \theta) \tag{1}$$

over the time interval $[t_0, t_1]$. Here $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{f}$ is a neural network with parameter vector (weights) θ. You will also study an application of an NDE to data from a simple linear differential equation.

The integral form of (1) is given by

$$\mathbf{z}(t) = \mathbf{z}(t_0) + \int_{t_0}^{t_1} \mathbf{f}(\mathbf{z}, \theta, t) dt \tag{2}$$

a.

Suppose we have data $\mathbf{z}(t_k), k = 1, \ldots, K$ of the solution. Formulate the loss function L, based on the mean squared error, used for fitting the weights of the NDE.

b.

Argue that when to optimise the loss function L, we need derivatives of L to $\mathbf{z}$ and θ.

c.

Let $\mathbf{a}(t) = \partial L / \partial \mathbf{z}(t)$. Using the chain rule,

$$\frac{\partial L}{\partial \mathbf{z}(t)} = \frac{\partial L}{\partial \mathbf{z}(t+\epsilon)} \frac{\partial \mathbf{z}(t+\epsilon)}{\partial \mathbf{z}(t)} \tag{3}$$

and (2), show that $\mathbf{a}(t)$ can be determined from the equation

$$\frac{d\mathbf{a}}{dt} = -\mathbf{a}^T \frac{\partial \mathbf{f}(\mathbf{z}, t, \theta)}{\partial \mathbf{z}} \tag{4}$$

where the subscript T indicates transpose.

d.

Describe a numerical solution procedure to determine $\mathbf{a}(t_0)$.

e.

To determine $\partial L / \partial \theta$, the augmented vector $\mathbf{x} = (\mathbf{z}, \theta, t)$ is used and the augmented equations are written as

$$\frac{d\mathbf{x}}{dt} = \mathbf{g}(\mathbf{z}, \theta, t) \tag{5}$$

where $\mathbf{g} = (\mathbf{f}, \mathbf{0}, 1)^T$. One introduces also $\mathbf{b} = (\mathbf{a}, \mathbf{a}_\theta, \mathbf{a}_t)$, where $\mathbf{a}_\theta = \partial L/\partial\theta$ and $\mathbf{a}_t = \partial L/\partial t$.

Show that

$$\mathbf{a}_\theta(t_0) = -\int_{t_1}^{t_0} \mathbf{a}^T \frac{\partial \mathbf{f}(\mathbf{z}, t, \theta)}{\partial \theta} dt \tag{6}$$

and describe the connection with backward propagation (as used in FNNs).

f.

In the notebook, first data from the linear differential equation

$$\frac{d\mathbf{z}}{dt} = \begin{pmatrix} -0.1 & -1.0 \\ 1.0 & -0.1 \end{pmatrix} \mathbf{z} \tag{7}$$

is used to train 'weights' from a 2×2 matrix. Determine how the loss function decreases versus the number of epochs for three different values of the number of data points $K = 75, 100, 200$. Use evweytime $lr, n\_epochs = [0.01, 1000]$.

g.

Also a more complicated ODE is formulated in the notebook (TestODEF and you are welcome to enter your own). Use $K = 200$ and implement for $\mathbf{f}$ a FNN with 1 hidden layer and 16 neurons to solve the NDE in this case. Plot the trajectory which results after the loss function has sufficiently decreased. Use this time $lr = 0.001$ and estimate the number of epochs necessary for the loss function to decrease below a reasonable trashelhold.