Introduction to Machine Learning for Geosciences
Homework 5

October 11, 2021

1 Theory

Consider a simple neural network with a single neuron in a single hidden layer, a scalar input x, training label y, a scalar output $\hat{y}$, and the squared error loss $\mathcal{L} = \frac{1}{2}(\hat{y} - y)^2$.

The input x is related to $\hat{y}$ through the function $\hat{y} = \sigma(w_2\sigma(w_1x + b_1) + b_2)$ where σ is the sigmoid activation function $\sigma(x) = \frac{1}{1+\exp(-x)}$. The 'learned' parameters of this network are $\{w_1, w_2, b_1, b_2\}$.

1. Using the chain rule, compute the gradient of the loss w.r.t the parameters $\nabla_{\boldsymbol{\theta}}\mathcal{L}$ where $\boldsymbol{\theta} = \{w_1, w_2, b_1, b_2\}$. It is easiest if you write out the chain rule terms in a column vector ' as partial derivatives first and then determine what each partial derivative should be.

2 Application

The file year co2 temp.csv contains annual average global CO2 emission and temperature change data for years 1961-2010.

1. Fit several orders of polynomials to the average global CO2 emissions as a function of year from the year 1961 to 1999. What order polynomial best predicts the data from 2000 to 2010?

2. Repeat exercise 1 by training a neural network to learn the regression. Comment on the ability of the network to generalize to unseen data from the year 2000 to 2010.

3. Train the neural network with CO2 data from a random selection of years including 2000 to 2010. Split the data as about 80% train and 20% test. Comment on the network's ability to generalize to test data now.