Exercise 1: Stochastic Gradient Descent

The file ex1data1.txt contains the dataset for our linear regression problem. The first column is the population of a city and the second column is the profit of a food restaurant in that city. A negative value for profit indicates that the restaurant does not do profit but looses money. Use this data to fit a linear model: \( h(x) = b + mx \). As shown in the lecture you can replace \( b \) and \( m \) with \( w_0 \) and \( w_1 \). To learn \( w_0 \) and \( w_1 \) you should use your own implementation of Stochastic Gradient Descent. Make sure you update \( w_0 \) and \( w_1 \) simultaneously (you can initialize both with 1). Run the process with 1500 iterations. Use 0.5 for learning rate.

Exercise 2: Further on Gradient Descent

What is the difference between batch gradient and stochastic gradient descent? And what is mini-batch gradient descent?

Exercise 3: Deep Learning

What is the fundamental difference between deep learning and traditional machine learning?

Exercise 4: Continue on Deep Learning

Draw a feedforward-neural network with two inputs \((x_1, x_2)\), 5 neurons on the first hidden layer, 3 neuron on the second hidden layer. The network should perform a binary classification.

Exercise 5: Softmax

What is softmax? And what do we use it for?

Exercise 6: Activation Functions

Name 3 activation functions and sketch their behaviour graphically. Why do we use activation functions?

\[ \text{http://www.is.inf.uni-due.de/courses/im_ws18/uebung/ex1data1.txt} \]
Exercise 7: RNN and LSTM

What is the advantage of RNN/LSTM over simple feedforward networks (simple neurons)?