HW #5

1. Bishop 5.26

2. Bishop 5.28

3. Exploring gradient descent

Use the Jupyter notebook mentioned on p. 19 of Mehta et al. (section IV F) to reproduce Fig. 9. Experiment with minimization of Beale's function initial conditions and hyperparameters and describe your insights. Test all 5 methods shown in Fig. 9, and attach the plots to your work.

4. NN for classification

Implement a two-layer (single hidden layer) feed-forward NN for $K=2$ classification. Train the NN on the Sonar & Ionosphere Datasets from HW#4.

Implement backpropagation of the first derivative of the error function and choose a gradient descent algorithm (either justify your choice or explore several strategies). Experiment with M, the number of nodes in the hidden layer, and report best classification accuracy of the trained NN on the entire dataset (do not forget to divide the data into training and test sets during NN training!)

Have you been able to do better than SVM & RVM? in HW#4

$$\begin{cases} z, & z \geq 0 \\ e^z - 1, & z < 0 \end{cases}$$

↓

ELU

Note: try $\tanh(z)$ and ELU activation functions and compare the results.