Artificial Neural Networks

Prof. Dr. Sen Cheng

Oct 14, 2019

Problem Set 2: Optimization

Tutors: Filippos Panagiotou (filippos.panagiotou@ini.rub.de), Mohammad M. Nejad (m.mohagheghi@ini.rub.de)

1. **Analytical minimization of a function**

   Study the minima of the function
   $$L(x) = x^4 - 4x^2 + 4 \tag{1}$$
   by following the steps below.

   (a) Find the roots of $\frac{\mathrm{d}}{\mathrm{d}x}L(x)$.

   (b) Determine whether the roots of the derivative are minima or maxima of $L(x)$. What is the sufficient condition for minima and maxima?

   (c) Plot L against x in Python for $x \in [-2, 2]$ and mark the minima (in red) and maxima (in green) in the plot.

2. **Numerical minimization of a function**

   Use gradient descent to find the minima of $L(x)$ from the problem above.

   (a) Implement gradient descent to find the minima of $L(x)$.

   (b) Starting the estimate from different values in $[-2, 2]$, run the gradient descent algorithm with $\eta = 0.01$ and $N = 50$ (N, number of iterations). Observe how the starting value determines which local minimum is found.

   (c) Set your initial guess to 1 and $N = 100$, change η linearly in $[0.05, 0.15]$ (`np.linspace`) and for each η estimate the minimum. Find the precision of the estimation by computing the absolute value of the difference between the estimated and the actual minimum for each η. Plot the differences against η. How does η affect the precision of the estimation?

   (d) Now set your initial guess to 1 and $\eta = 0.01$ and change N linearly in $[1, 100]$. How does N affect the precision of the estimation?

3. **The role of the learning rate**

   To better understand the implications of choosing an appropriate learning rate η in gradient descent let us focus on a simpler loss function $V(x) = x^2$ because it has only one minimum.

   (a) Use gradient descent to find the minimum of this function. To better understand the method you should keep track of the estimated minimum in each iteration, find the corresponding point on the function $V(x)$ and plot them together with the loss function.

   (b) Set your initial guess to 1 and try few different η in $[0.1, 1.1]$ and observe the changes in the estimate. Find at least one η for which the estimate converges and one η for which it diverges.