Big Data in Experimental Physics: NN

TAs

August 23, 2020

1   Back propagation

As `README.md` has explained, the computation process is

$$v^{(1)} = W^{(1)}v^{(0)}$$
$$v^{(2)} = \text{ReLu}(v^{(1)})$$
$$v^{(3)} = W^{(2)}v^{(2)}$$

The loss is computed in the following way

$$E = \frac{1}{2}\sum_i (v_i^{(3)} - y_i)^2$$

To minimize the loss, we can use derivative to update the weights as E is a function of all weights. The gradient for all the weights can be obtained in a way similar to the following process.

$$\frac{\partial E}{\partial v^{(3)_i}} = v_i^{(3)} - y_i$$
$$\frac{\partial E}{\partial W_{i,j}^{(2)}} = \frac{\partial E}{\partial v^{(3)_i}}\frac{\partial v_i^{(3)}}{\partial W_{i,j}^{(2)}}$$
$$= \frac{\partial E}{\partial v^{(3)_i}}v_j^{(2)}$$
$$\frac{\partial E}{\partial v_j^{(2)}} = \sum_i \frac{\partial E}{\partial v_i^{(3)}}\frac{\partial v_i^{(3)}}{\partial v_j^{(2)}}$$
$$= \sum_i \frac{\partial E}{\partial v_i^{(3)}}W_{i,j}^{(2)}$$

As for the ReLU function,

$$\frac{\partial v_i^{(2)}}{\partial v_i^{(1)}} = \begin{cases} 1 & v_i^{(1)} > 0 \\ 0 & v_i^{(1)} \leq 0 \end{cases}$$

The second half of the problem is to compute

$$\text{argmax}_{i,j}\frac{\partial E}{\partial W_{i,j}^{(1)}}$$