Machine Learning Worksheet 08

Neural Networks 2

A simple neural network has as loss function

$$E(\boldsymbol{w}) := \frac{1}{m}\sum_{i=1}^{m} f(z_i - \boldsymbol{w}\cdot\boldsymbol{x}_i) + \lambda\|\boldsymbol{w}\|^2/2$$

where

$$f(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| < 1 \\ |x| - \frac{1}{2} & \text{otherwise.} \end{cases}$$

In these, $x \in \mathbb{R}^d$ are the data, $w \in \mathbb{R}^d$ are the weights, and $z \in \mathbb{R}$ the target outputs for m data points. The λ is a constant.

Problem 1: Compute the gradient of $E(w)$ w.r.t. w, when optimising over all data.

Define $\boldsymbol{g}_i$ as the gradient of the ith data point.

$$\boldsymbol{g}_i = \begin{cases} (\boldsymbol{w}\cdot\boldsymbol{x}_i - z_i)\boldsymbol{x}_i & \text{if } |\boldsymbol{w}\cdot\boldsymbol{x}_i - z_i| < 1 \\ \text{sgn}(\boldsymbol{w}\cdot\boldsymbol{x}_i - z_i)\boldsymbol{x}_i & \text{otherwise} \end{cases}$$

then

$$\frac{\partial E}{\partial \boldsymbol{w}} = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{g}_i + \lambda\boldsymbol{w}$$

Problem 2: How do you minimise $E(\boldsymbol{w})$? Write down the equation of a method to minimise the loss for a specific training instance $(\boldsymbol{x}_i, z_i)$.

As formulated, there are many solutions. This may be one of the simpler ones.

Using the above definition of $\frac{\partial E}{\partial \boldsymbol{w}}$, we can write

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta\frac{\partial E}{\partial \boldsymbol{w}}$$

which leads to

$$\boldsymbol{w} \leftarrow (1 - \lambda\eta)\boldsymbol{w} - \eta\boldsymbol{g}_i$$

Problem 3: In the uploaded Jupyter notebook neuralnetworks1.ipynb, implement the momentum in both neural network implementations. See that learning with the momentum can be considerably faster than without.

Plot several (at least 6) training curves with and without momentum, while experimenting with different values for the learning rate and the momentum. Submit the plots, as well as the relevant lines of your code changes.

Done.