Neural networks: Assignment 1

February 28, 2024

Please submit your assignments carefully written down in LaTeX or jupyter notebook, as applicable.

Problem 1: Simulate a boolean function (2 points)

Write a neural network NN with four inputs and one output such that

$$\mathrm{NN}(x_1, x_2, x_3, x_4) = 1 + x_1x_2 + x_2x_3 + x_3x_4 \pmod 2$$

for every $x_1, x_2, x_3, x_4 \in \{0, 1\}$.

You can use any architecture and any activation functions that you like.

Problem 2: Equivalence of activations (3 points)

(a) Recall the step function $H(x) = \mathbb{1}(x \geq 0)$ and the sigmoid $S(x) = \frac{1}{1+\exp(-x)}$. For every $\varepsilon > 0$, describe a neural network NN with one input, one output and sigmoid activations such that

$$|\mathrm{NN}(x) - H(x)| \leq \varepsilon$$

for every x such that $|x| > \varepsilon$.

(b) Let $f : [-1, 1] \to \mathbb{R}$ be a continuous function and $\varepsilon > 0$. Describe a neural network NN with one input, one output, using step and identity activations, such that for every $-1 \leq x \leq 1$

$$|\mathrm{NN}(x) - f(x)| \leq \varepsilon \ .$$

Hint: Note that for $b_1 \leq b_2$, we have $H(x - b_1) - H(x - b_2) = \mathbb{1}(b_1 \leq x < b_2)$.

Problem 3: Gradients for one neuron (3 points)

Let $x, w \in \mathbb{R}^d, b \in \mathbb{R}$.

a) Let $f_R(x) = \mathrm{ReLU}(w^T \cdot x + b)$ be a ReLU neuron and $C_{\mathrm{sq}}(y, \overline{y}) = (y - \overline{y})^2$ denote the square loss for $\overline{y} \in \mathbb{R}$. Calculate $\frac{\partial}{\partial w_i} C_{\mathrm{sq}}(f_R(x), \overline{y})$.

b) Let $f_S = S(w^T \cdot x + b)$ be a sigmoid neuron and $C_{\mathrm{cross}}(y, \overline{y}) = -\overline{y} \ln y - (1 - \overline{y}) \ln(1 - y)$ be the *cross-entropy* loss for $0 \leq \overline{y} \leq 1$. Calculate $\frac{\partial}{\partial w_i} C_{\mathrm{cross}}(f_S(x), \overline{y})$.

c) Let $f_T = \tanh(w^T \cdot x + b)$ be a neuron with activation $\tanh(x) = \frac{\exp(2x)-1}{\exp(2x)+1}$ and $C_{\mathrm{hinge}}(y, \overline{y}) = \max(0, 1 - y\overline{y})$ be the *hinge loss* for $\overline{y} \in \{-1, 1\}$. Calculate $\frac{\partial}{\partial w_i} C_{\mathrm{hinge}}(f_T(x), \overline{y})$.

Problem 4: Accuracy at initialization (1 point)

Assume there is a dataset with M points and labels $x^{(i)} \in \mathbb{R}^d$, $\overline{y}^{(i)} \in \{0, 1, 2, 3\}$ for $i = 1, \ldots, M$. What accuracy can you expect for this dataset at initialization?

Problem 5: Playing with code (5 points)

Take the code from the `network_first.ipynb` file (it will be updated on Thursday, but you can already fill the details yourself). Then, try at least one of the following ways to modify the code:

- Increase the number of neurons.
- Change the activation function (eg., ReLU) and/or the cost function (eg., cross-entropy or one of its variants).

- Add another layer to the network.
- Change the size of a mini-batch.
- Add L2 regularization.

You can also consider other changes and improvements. Report on your findings.