6.390 Introduction to Machine Learning

Recitation Week #7

Issued October 17, 2022

1. Nori thinks about reLU units and wonders whether there is a better alternative for an activation function, and decides to explore the LUre function, defined as:

$$f_{\mathrm{LUre}}(z) = \min(z, 0).$$

(a) Sketch $f_{\mathrm{reLU}}(z)$ and $f_{\mathrm{LUre}}(z)$.

(b) What is the derivative of this function, $\frac{df_{\mathrm{LUre}}(z)}{dz}$?

(c) Nori's friend Ori thinks this is cool and suggests making a neural network with *two* activation functions per layer, in particular,

$$a^l = f_{\mathrm{LUre}}(f_{\mathrm{reLU}}(z^l)).$$

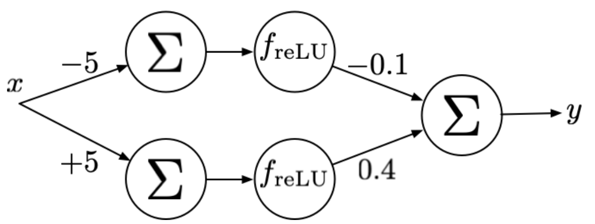
Explain what effect this will have on the network.

(d) Nori's other friend Dori thinks we should try this trick with two reLUs, so that,

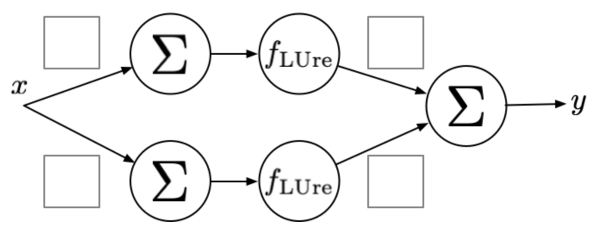
$$a^l = f_{\text{reLU}}(f_{\text{reLU}}(z^l)).$$

Explain what effect this will have on the network.

(e) Nori finds a neural network trained by her nemesis Smori that takes a single-dimensional input ($d = 1$) and looks like this (with no constant offsets into the summation):



She sees that it computes $\hat{y} = -0.1 f_{\text{reLU}}(-5x) + 0.4 f_{\text{reLU}}(5x)$ and is very curious to see if she can replace those reLU activation units with her own LUre's. Please help her find another neural network that computes exactly the same function as the one above (that is, maps any input x to the same output as the original one). Provide a set of weights that achieves this in the boxes on the diagram below.



(f) Is the resulting set of weights found by Nori in (d) unique? Would it be possible to create an equivalent network if Smori were to include offsets at each neuron? Why or why not?

2. We want to design a neural network to solve each of these problems. For each one, indicate a good choice of:

   - representation of actual value, $y^{(i)}$, for each of the $i = 1, \cdots, n$ data points in the training data set,
   - activation function on the output layer,
   - loss function for training, $\mathcal{L}(g^{(i)}, y^{(i)})$, where $g^{(i)}$ is the guess, and $y^{(i)}$ is the actual label, for each of the $i = 1, \cdots, n$ data points.

   Note: Write a mathematical expression for the loss function, not just the type of loss in words. You can assume $g^{(i)}$ (outputs after activation function is applied in the last layer) and $y^{(i)}$ are scalars or vectors depending on context. Assume that the objective function used for training the neural network takes the form:

$$J(W) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(NN(x^{(i)}; W), y^{(i)}),$$

   and you are asked to find the expression for $\mathcal{L}(g^{(i)}, y^{(i)})$, where the guess for each input data point, $g^{(i)} = NN(x^{(i)}; W)$, is the output of your proposed neural network with weights W.

   (a) Predict when a train will arrive based on the day, time, and weather, in minutes relative to the scheduled arrival time. If your prediction is after the train actually arrives, it has loss 100. If before, then the loss is the number of minutes early you predict.

      i. Representation of target output value:
         - ○ thermometer-encoding
         - ○ real number
         - ○ one-hot vector
         - ○ any vector of values in $\{0, 1\}$

      ii. Output activation function:

      iii. Assume that both guess and actual are represented in terms of minutes. Loss function (provide full equation):
      $\mathcal{L}(g^{(i)}, y^{(i)}) =$

(b) Predict which items-out of 5 possible items sold by Things 'R' Us–a shopper will purchase during one shopping trip, based on their previous shopping history. You care only about whether or not an item is bought (not the quantity purchased), and any given customer can order multiple items.

i. Representation of target output value:
- ◯ thermometer-encoding
- ◯ real number
- ◯ one-hot vector
- ◯ any vector of values in $\{0, 1\}$

ii. Output activation function:

iii. Loss function (provide full equation):
$\mathcal{L}(g^{(i)}, y^{(i)}) =$

(c) Predict the profession of a person (out of 10 possible values) based on their walking speed and clothing. Assume that each person only has a single profession.

i. Representation of target output value:
- ◯ thermometer-encoding
- ◯ real number
- ◯ one-hot vector
- ◯ any vector of values in $\{0, 1\}$

ii. Output activation function:

iii. Loss function (provide full equation):
$\mathcal{L}(g^{(i)}, y^{(i)}) =$