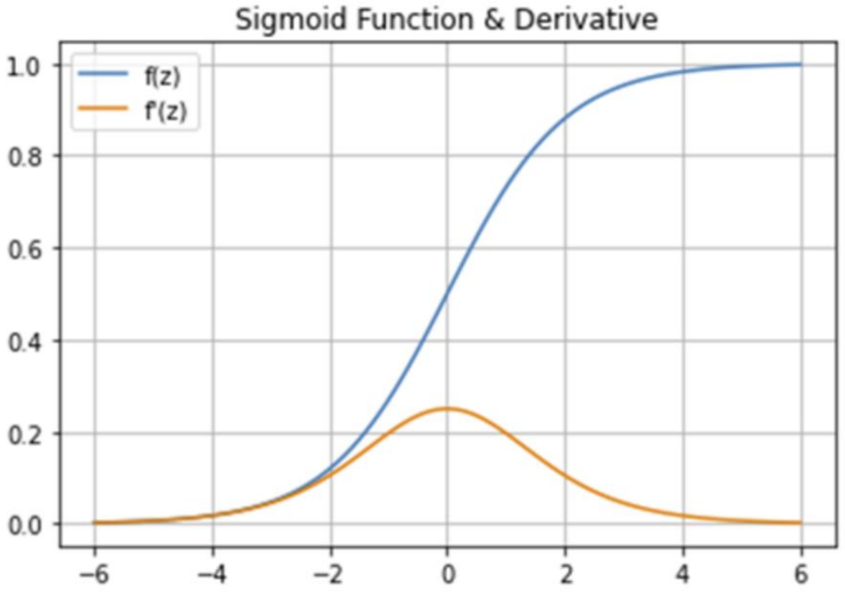
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Deep Learning for Physicists

Tutorial #1

Kosmas Kepesidis

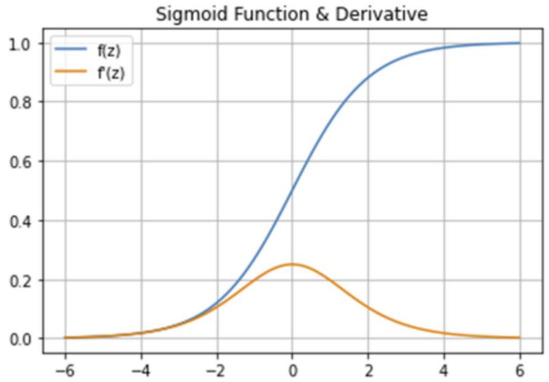
Question 1: Why was the logistic activation function a key ingredient in training the first artificial neural networks?



- Inspired by activations in biological neurons
- It is bounded between [0,1]
- It has a well-defined non-zero derivative everywhere
  - This is important for Gradient Descent. When neural networks were first theorized, people assumed an on-off behavior of neurons, i.e. step functions. The immediate problem with step functions was that there is no gradient to work with and Gradient Descent cannot “move” on flat subphases.
  - Logistic activations allow Gradient Descent to do some progress at every step.
- Disadvantages:
  - Vanishing gradients – low values of derivatives leading to non-significant update
  - Computationally expensive

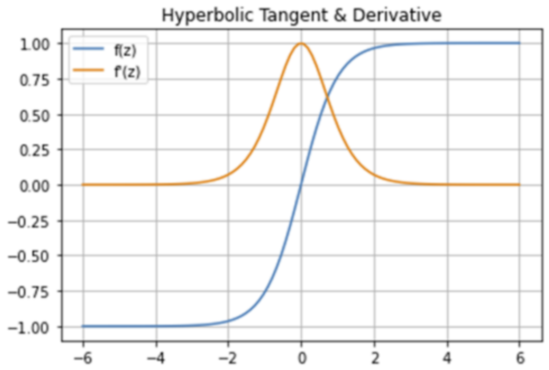
$$f(z) = \frac{1}{1+e^{-z}} \qquad f'(z) = \frac{e^{-z}}{(1+e^{-z})^2}$$

Question 2: Name and draw three popular activation functions. Draw also their first derivatives



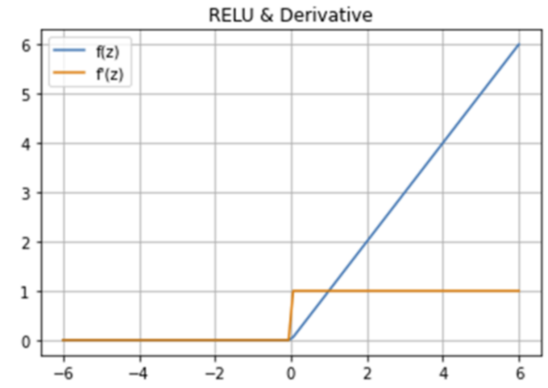
- Logistic (sigmoid) activation function

$$f(z) = \frac{1}{1+e^{-z}} \qquad f'(z) = \frac{e^{-z}}{(1+e^{-z})^2}$$



- Hyperbolic Tangent (Tanh)

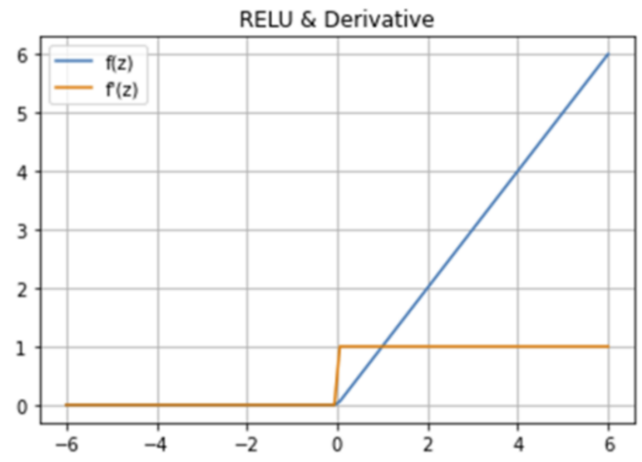
$$f(x) = \tanh(z) \qquad f'(x) = \mathrm{sech}(z)$$



- Rectified Linear Unit (ReLU)

$$f(x) = \max(0, z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases}$$

Question 3: Name and draw three popular activation functions. Draw also their first derivatives



- Rectified Linear Unit (ReLU)

  - Avoids the problem of vanishing gradients since its derivative if finite for $z > 0$

  - Computationally faster due to its simple form

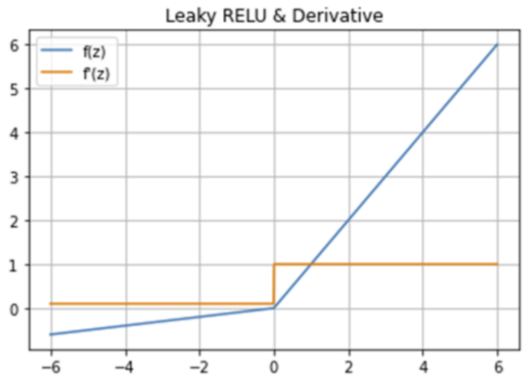
  - Disadvantage (Dying-ReLU problem):

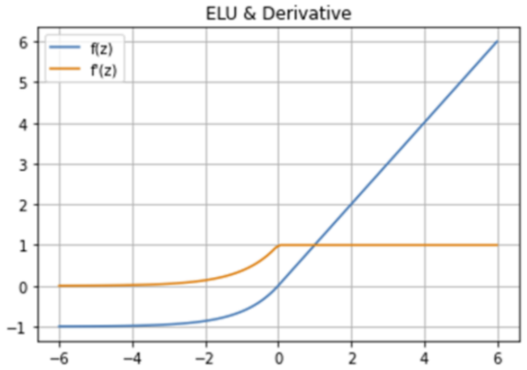
  For z<0, derivative is zero. This eventually stops the learning process, neuron stops responding to any changes. This leads to dead neurons which are not able to update the weights anymore in backpropagation.

  - Solutions:

  Modification of ReLU: Leaky ReLU, ELU, …

$$f(x) = \max(0, z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases}$$





Question 4: Explore the training of neural networks using the TensorFlow Playground

- Train the default network using default settings. How quickly does it find a good solution to all classification problems?
  - Converges fast for 3 out of 4 problems. The last one is of higher complexity and needs additional layers
- Change the activation function from Tanh to ReLU. Does the optimization converge faster? Also, what can you say about the decision boundaries that are generates by the two kinds of activation functions?
  - It converges even faster, but this time it produces linear boundaries due to the shape (linearity) of the ReLU function
- Try adding some additional layers and explore the types of patterns that are learned at each layer. What do you observe?
  - The more layers there are, the more complex patterns are learned