A Gentle Introduction to Deep Learning

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

January 2021

Outline

- Learning Goals
- Deep Learning Intuition
- Commonly Used Activations
- Summary

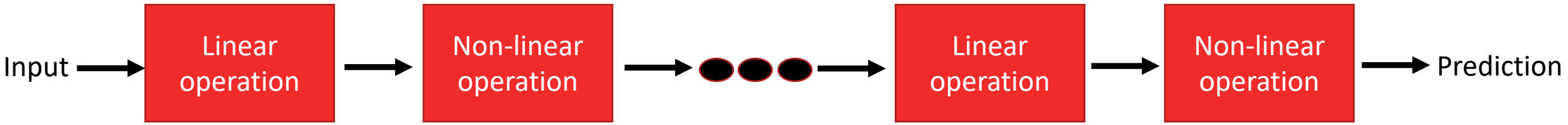
Learning Goals

- Get an intuitive feel for how the majority of deep learning methods work

- Get familiar with different activation functions used in deep learning

Deep Learning Intuition

- Alternated stack of linear and non-linear operations
- Non-linear operations that come immediately after a linear operation are called “activations”
- The activation at the end of the network determines if the model is a regression or classification network
- You can have two consecutive non-linear operations
- Two consecutive linear operations often do not make sense



UNIVERSITY OF CALGARY

Deep Learning Intuition

C = A × B → Equivalent linear operation

$$Y = A \times B \times X$$

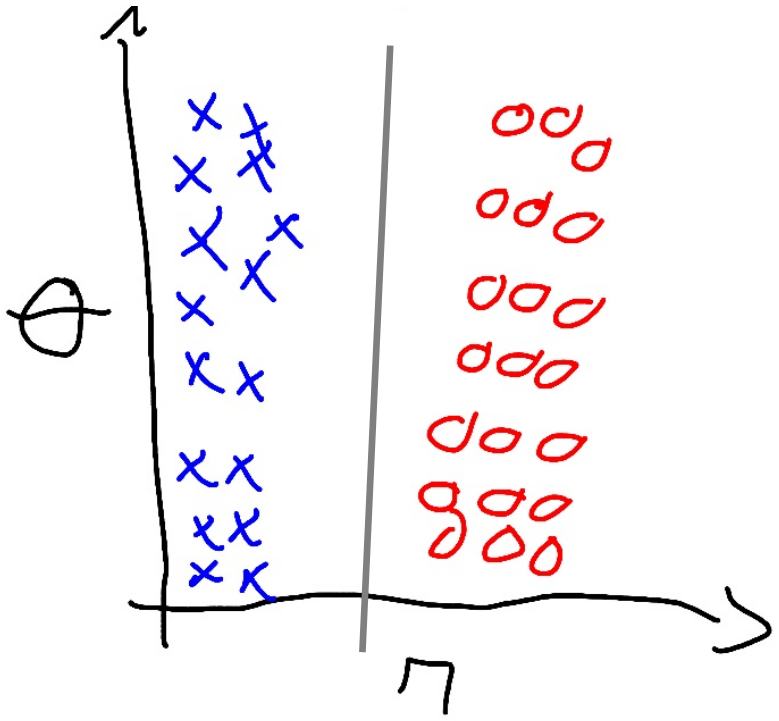
X → Input

Y → Output

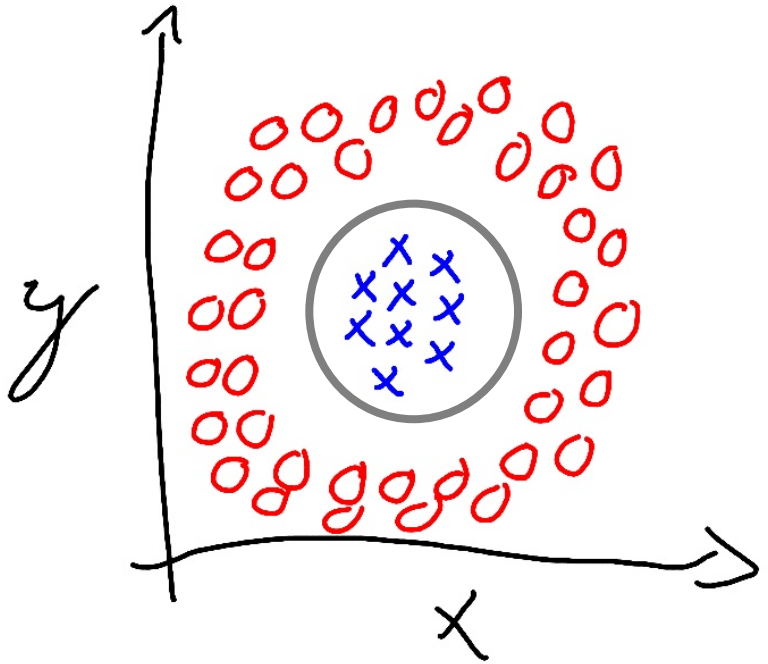
A → Linear operation

B → Linear Operation

Deep Learning Intuition



Linear model



Non-linear models allow you to get more complex decision boundaries.

UNIVERSITY OF CALGARY

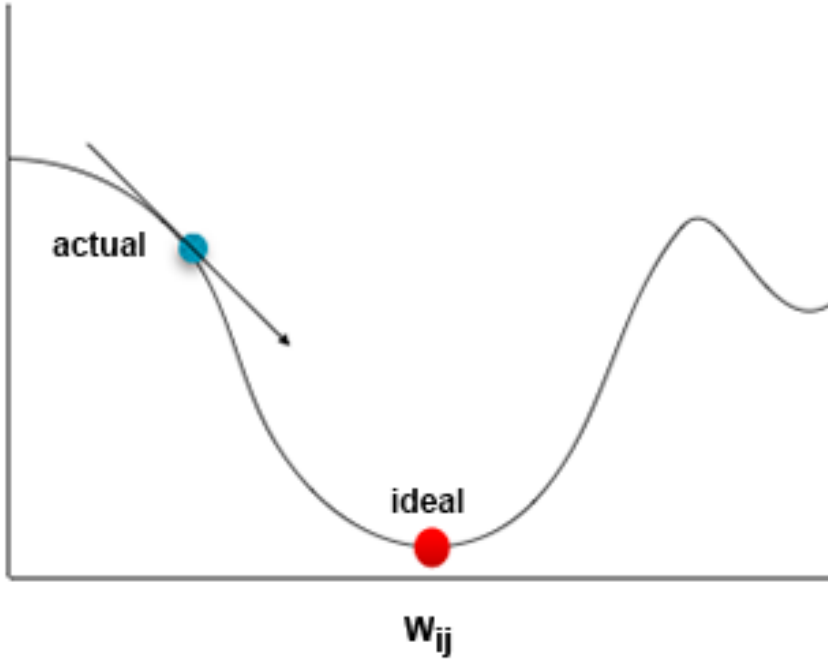
Deep Learning Intuition

1. Data
2. Model
3. Cost function or loss or objective

Fit the data to the model by minimizing your cost function.

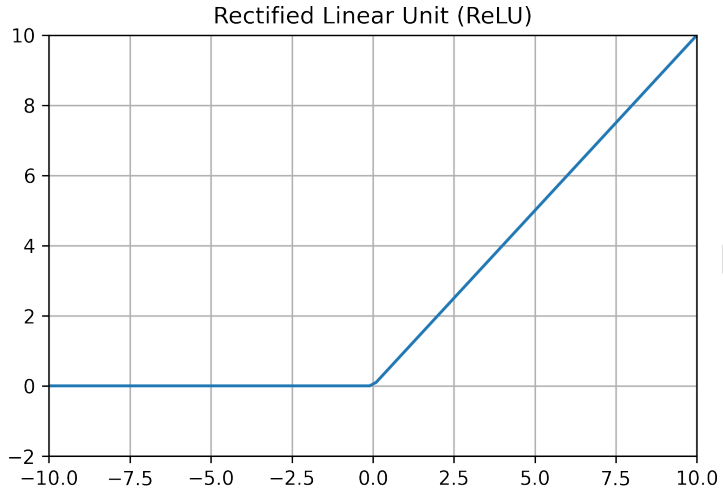
Deep Learning Intuition

- Gradient descent optimization of the cost function
    - Linear and non-linear operations need to be differentiable
- Compute the gradient across the training set (the whole set or mini-batches)
- Update the model weights by giving a step in the opposite direction (i.e., minimize the cost)
- Compute the average cost function in the train and validation sets after each epoch

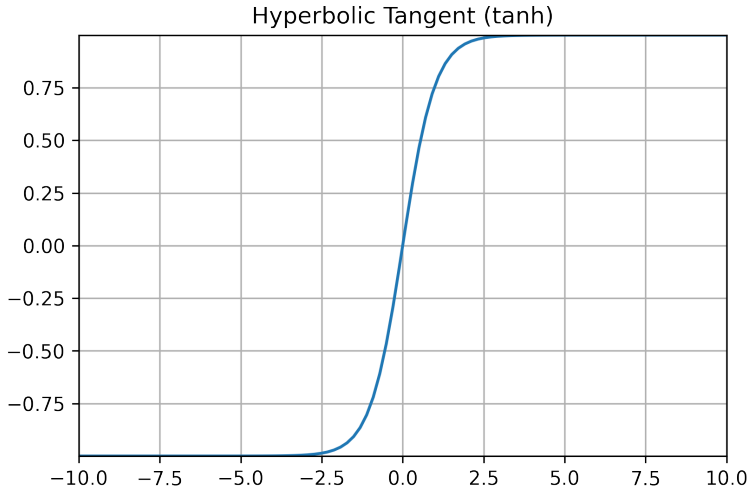


UNIVERSITY OF CALGARY

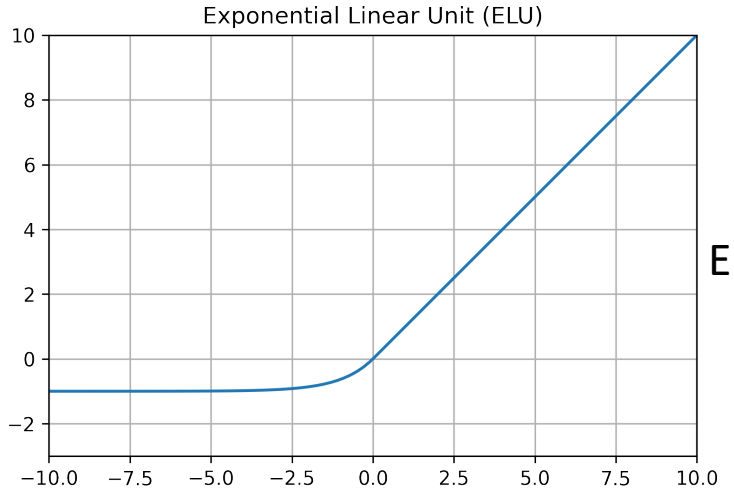
Activations



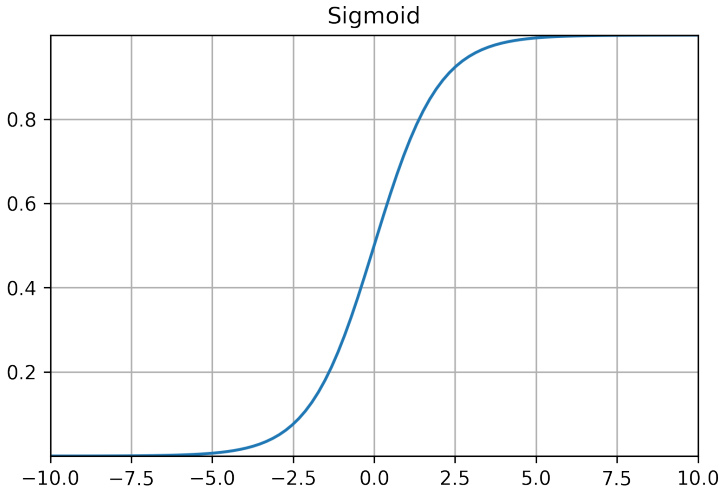
$$\text{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$



$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$$



$$\text{ELU}(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^{x}-1) & x < 0 \end{cases}$$

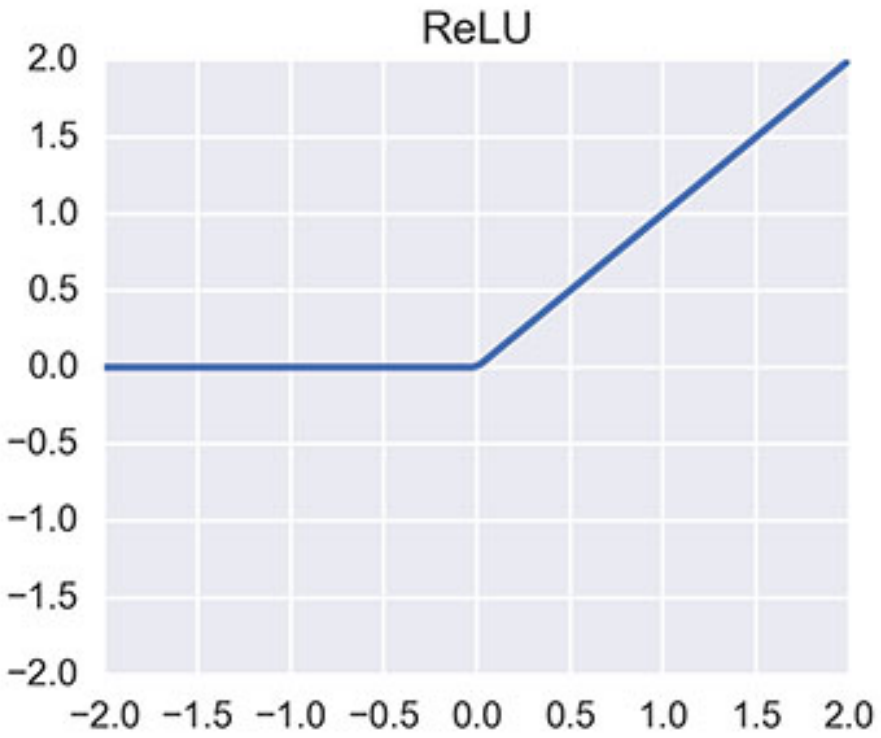


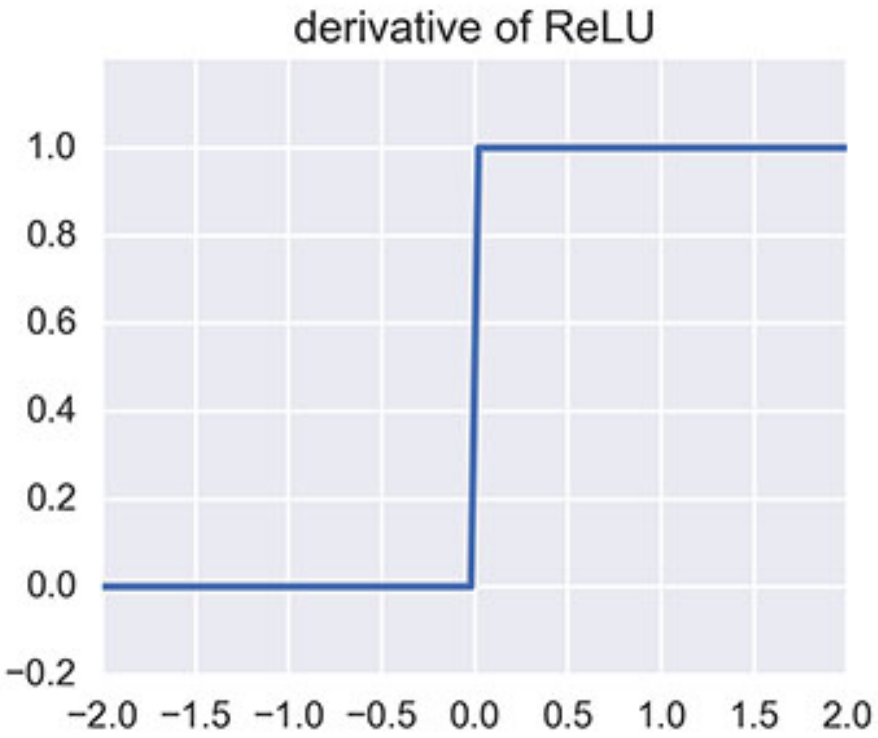
$$sigmoid(x) = \frac{1}{1+e^{-x}}$$

https://keras.io/api/layers/activations/

UNIVERSITY OF CALGARY

Activations - ReLU





$$\text{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$\frac{dReLU(x)}{dx} = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

Activations - Softmax

$$softmax(\vec{z}) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

- Softmax converts a real vector to a vector of categorical probabilities.
- The elements of the output vector are in range (0, 1) and sum to 1.
- Softmax is often used as the activation for the last layer of a classification network -> results are interpreted as a probability distribution.

Summary

- Deep learning models alternate between differentiable linear and non-linear operations
- Deep learning models are fit (i.e., trained) to the data by minimizing a cost function using gradient descent methods
- There are many potential non-linear operations
- ReLUs are commonly used due to their computational simplicity and simple derivative

Thank you!

UNIVERSITY OF CALGARY