Machine Learning

Seminar 6

1. Explain the principle of the gradient descent algorithm. Accompany your explanation with a diagram. Explain the use of all the terms and constants that you introduce and comment on the range of values that they can take.

2. For the following activation functions, write down the equation, function input range, function output range, draw a figure of input and output, and **write some codes to implement them as functions in Python**.

- Sigmoid
- Softmax
- Hyperbolic Tangent (tanh)
- ReLU
- Leaky ReLU(LReLU)
- Exponential Linear Unit (ELU)
- Step Function
- Swish

3. Where to use and which type of Activation Functions?

- For regression problems (Only 1 neuron, multiple inputs, real-world outputs), a linear activation function must be used.
- For multi-class classification problems, use Softmax at the output layer
- For multi-label and binary classification problems, use the Sigmoid activation function.
- Sigmoid and hyperbolic tangent activation functions must be never used in the hidden layers as they can lead to vanishing gradients.
- For networks where unnecessary neurons need to turn OFF, use ReLU as the activation function because it also works as a dropout layer. In case there is confusion about which activation function, use ReLU.It is used in most CNN problems.
- For deep neural networks having greater than 40 layers, use the swish activation function.