**Machine Learning in Finance - Homework 5 – Due 10.08.2024**

Textbook reading:

Chapters 5 starting page 129

1. **for a range of inputs from -infinity to +infinity,  what are the output ranges for the elu, exponential, gelu, linear, and softmax activation functions in**

[**https://www.tensorflow.org/api\_docs/python/tf/keras/activations**](https://www.tensorflow.org/api_docs/python/tf/keras/activations)

* ReLU:

max(x, 0)

(0,

* Exponential Linear Unit (ELU):
* Gaussian Error Linear Unit (GELU):
* Linear (pass-through):
* Softmax:

It also sums to one.

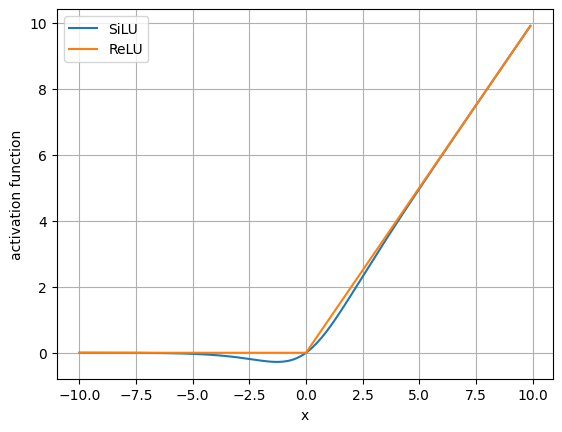
1. **What is regularization and why does it help create better models?**

Regularization adds a penalty term to the model's loss function, discouraging it from fitting the training data too closely by controlling the magnitude of model parameters.

On the bias-variance trade off, regularization reduces variance by adding bias.

1. **Using matplotlib plot the swish function formula. How is it different from 'relu'? What advantages does it have over relu? What disadvantages does it have?**

* Swish (SiLU):



ReLU has a kink at x=0. SiLU is differentiable everywhere, which has an advantage over ReLU in gradient calculations. On the other hand, the gradients of SiLU is slightly more computational costly.

1. **In Section5.2\_ANN.ipynb, replace 'relu' with 'swish' and replace the single layer model with a model containing two hidden layers, each containing 45 neurons.** 
   1. **Plot the loss history and the accuracy history. How do they compare to the same model using 'relu' instead of 'swish'?**

A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generated

Figure 1. Loss history of ReLU(left) vs. SiLU(right)

A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generated with medium confidence

Figure 2. Accuracy history of ReLU(left) vs. SiLU(right)

By changing the model to 2 layers with SiLU activation function and 45 neurons each layer, the model seems to have overfitted the training dataset.

* 1. **How many parameters are estimated in this model. Use the formulas to compute the number of parameters to verify that the input values match the actual number of parameters.**

Input layer: 784

Hidden layer #1: 45

Hidden layer #2: 45

Output layer: 10

Total Parameters

= (784 x 45 + 45) + (45 x 45 + 45) + (45 x 10 + 10)

= 35325 + 2070 + 460 = 37855

A screenshot of a computer

Description automatically generated

* 1. **Which layers in the model have no parameters to fit?**

Input layer.