# ECSE 552 – HW1 Report

Arnaud Ridard (McGill ID: 261091214), Saif Shahin (McGill ID: 260964749)

February 7th, 2023

## 1 Multi-Layer Perceptron Training from Scratch

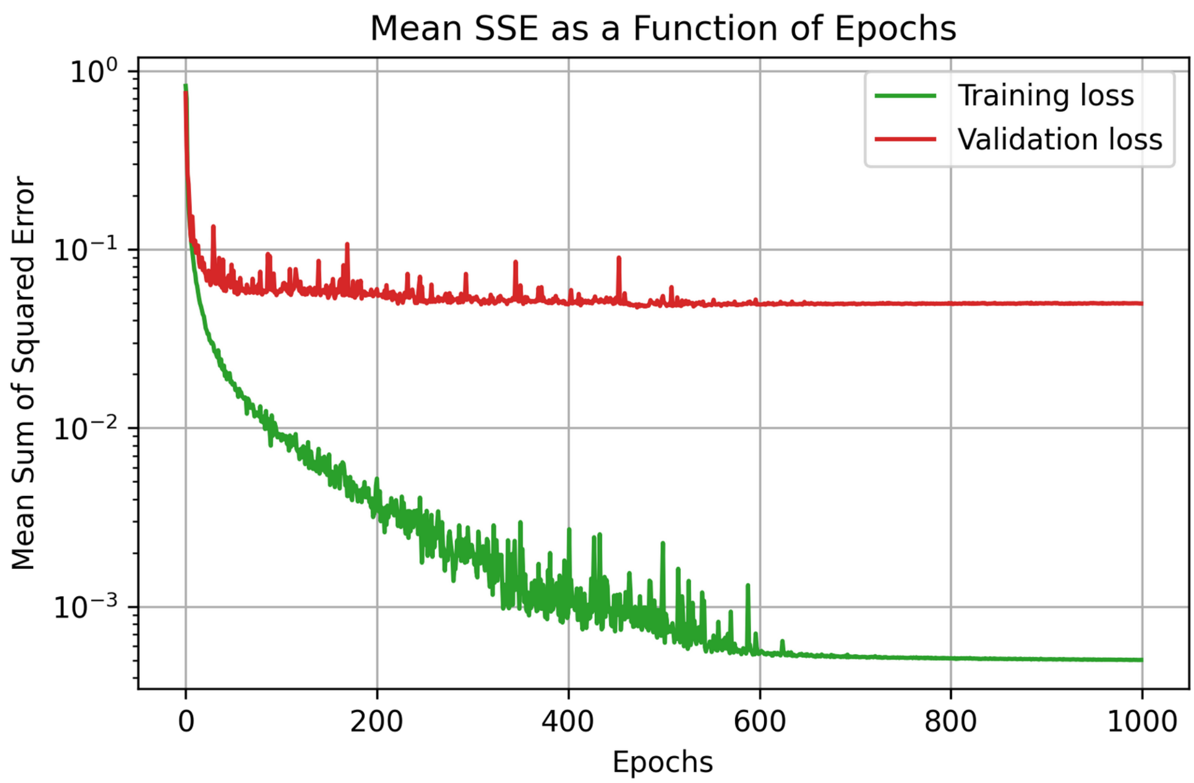


Figure 1: Training and validation loss (mean of SSE) for task 1.

## 2 Classification

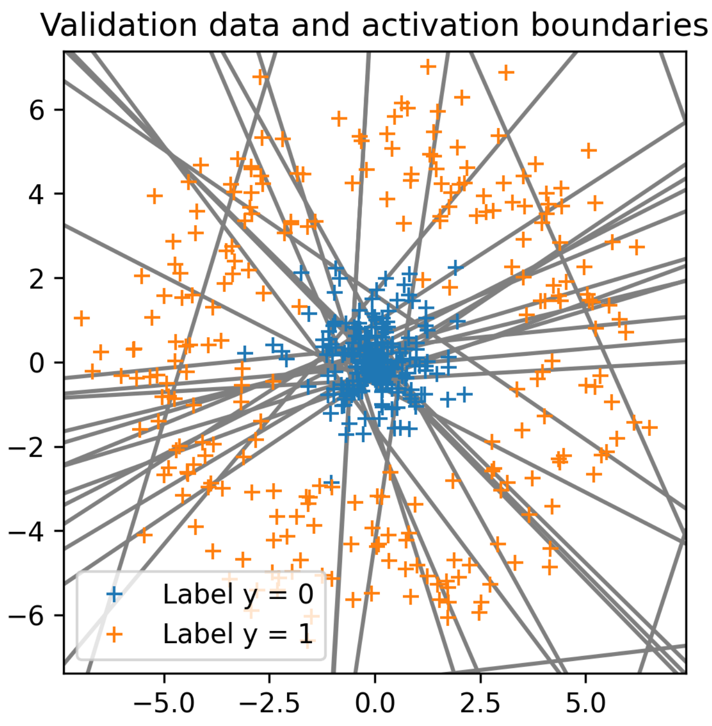


Figure 2: Validation data points and hidden units activation boundaries (30 hidden units).

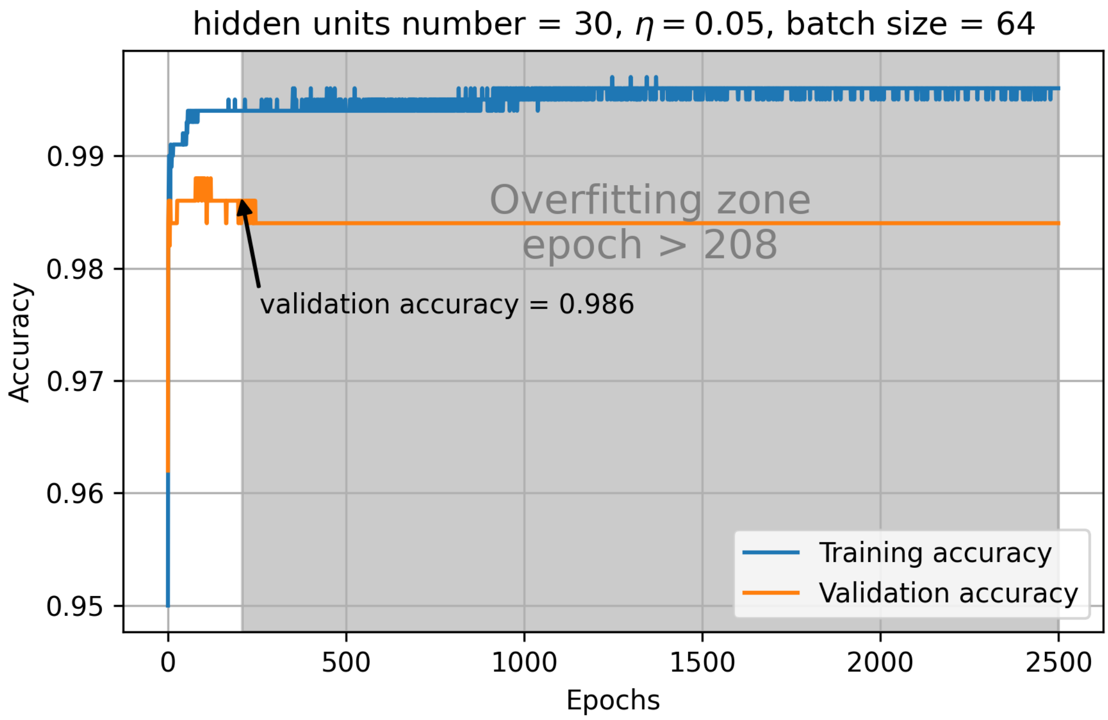


Figure 3: Training and validation binary classification accuracy for task 2.

## 3 Saturation of an output unit with sigmoid activation

We denote the cost function $J$ to be the negative log-likelihood

$$\begin{aligned} J(z) &= -\log[P(y|z)] \\ &= -\log\{1+\exp[-(2y-1)z]\}\,. \end{aligned}$$

Then the derivative of the cost function is

$$\frac{\mathrm{d}J}{\mathrm{d}z} = \frac{(2y-1)\exp[-(2y-1)z]}{1+\exp[-(2y-1)z]} = \frac{2y-1}{1+\exp[(2y-1)z]}$$

1. Suppose $z$ is a large positive real and $y=1$

$$\frac{\mathrm{d}J}{\mathrm{d}z} = \frac{1}{1+\exp z} \underset{z\to+\infty}{\longrightarrow} 0\,.$$

2. Suppose $z$ is a large positive real and $y=0$

$$\frac{\mathrm{d}J}{\mathrm{d}z} = \frac{-1}{1+\exp -z} \underset{z\to+\infty}{\longrightarrow} -1\,.$$

3. Suppose $z$ is a large negative real and $y=1$

$$\frac{\mathrm{d}J}{\mathrm{d}z} = \frac{1}{1+\exp z} \underset{z\to-\infty}{\longrightarrow} 1\,.$$

4. Suppose $z$ is a large negative real and $y=0$

$$\frac{\mathrm{d}J}{\mathrm{d}z} = \frac{-1}{1+\exp -z} \underset{z\to-\infty}{\longrightarrow} 0\,.$$