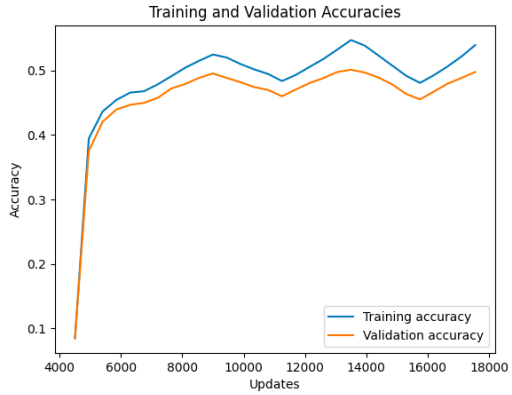
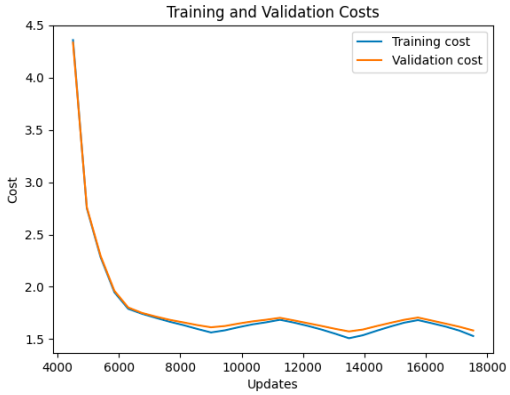
Assignment 3 – Report – Thomas Birchler

Gradient computation correctness

The following plots show the cost for one layer and two layers with each 50 nodes for 3 cycles without BN. These are very similar to the results I had in Assignment 2.





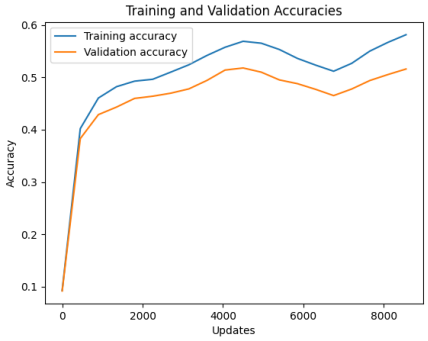
When comparing the gradients between the analytical and the numerical approach, my errors are all in the range of e-8 and e-9. This leads me to believe my gradient computation is correct. This is also for networks with several layers.

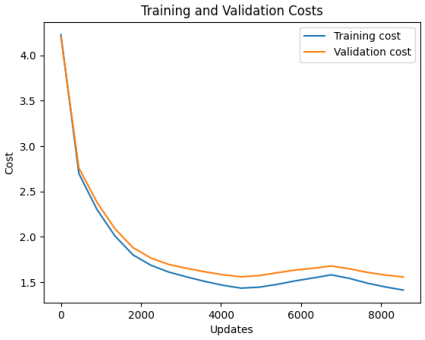
Without BN

3-layer Network without BN

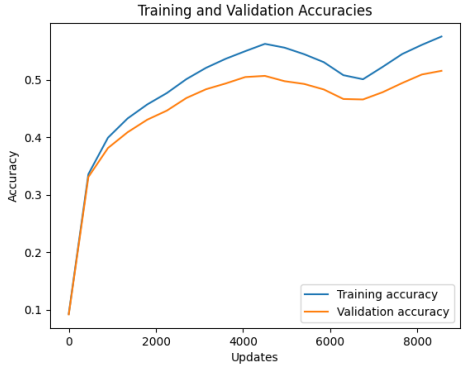
These two figures show the accuracy and cost of the training for 2 cycles.

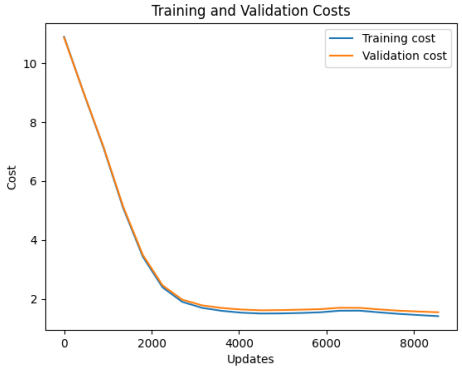
Lambda = 0.005, eta = [0.00001, 0.1], layers = [50, 50, 10], batch_size = 100, without BN, initialization: “He”. Final test accuracy: 0.533



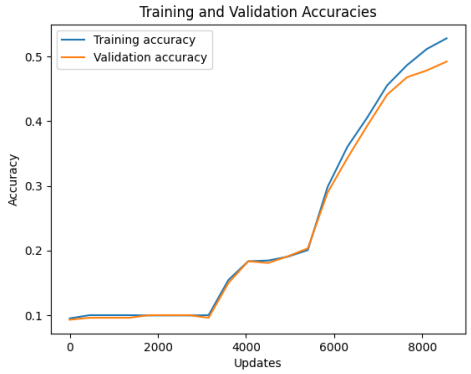


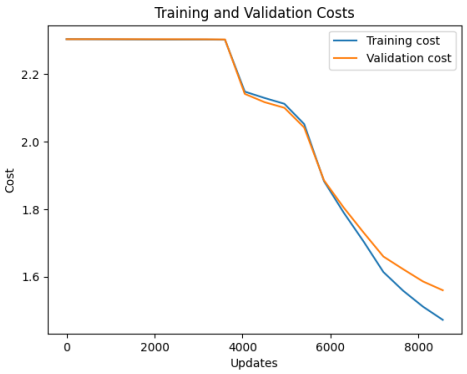
Initialization: sigma = 0.1, Final Test accuracy: 0.527



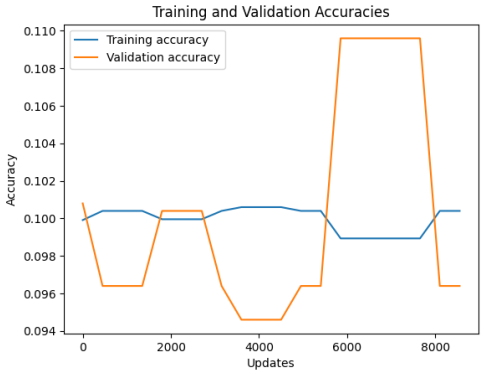


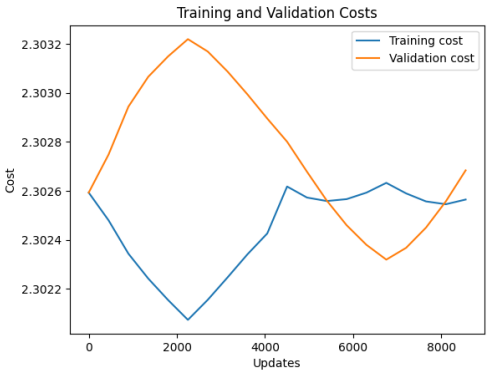
Initialization: sigma = 0.001, Final Test accuracy: 0.506





Initialization: sigma = 0.0001, Final Test accuracy: 0.1



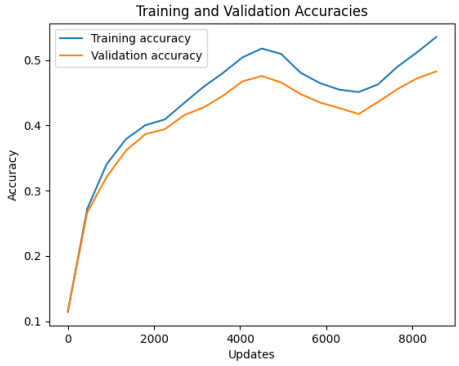


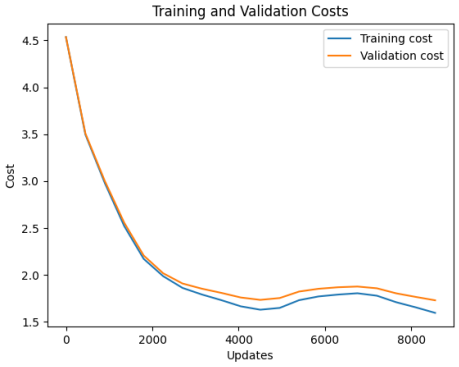
9-layer network without BN

The following plots show the results for the following settings:

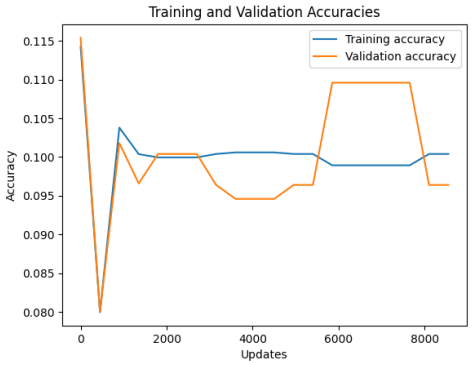
Lambda = 0.005, eta = [0.00001, 0.1], layers = [50, 30, 20, 20, 10, 10, 10, 10], batch_size = 100, initialization: “he”.

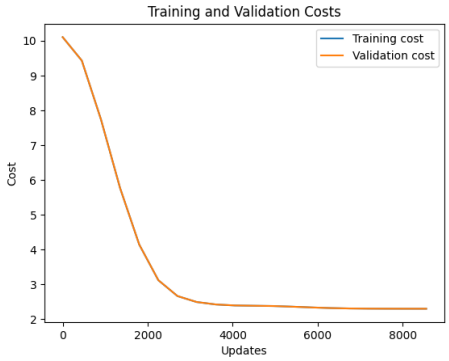
Final test accuracy: 0.500



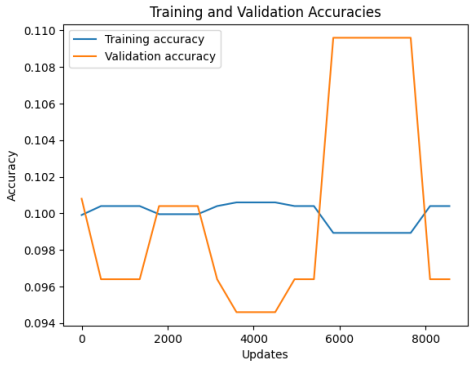


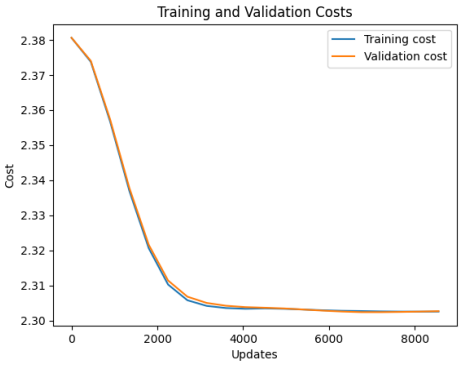
Initialization: sigma = 0.1, Final Test accuracy: 0.1



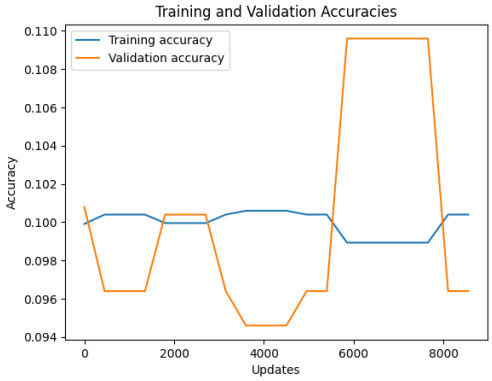


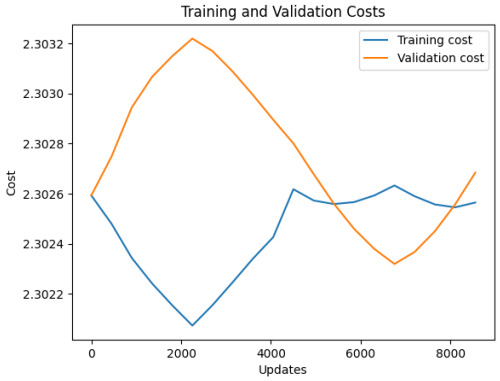
Initialization: sigma = 0.001, Final Test accuracy: 0.1





Initialization: sigma = 0.0001, Final Test accuracy: 0.1



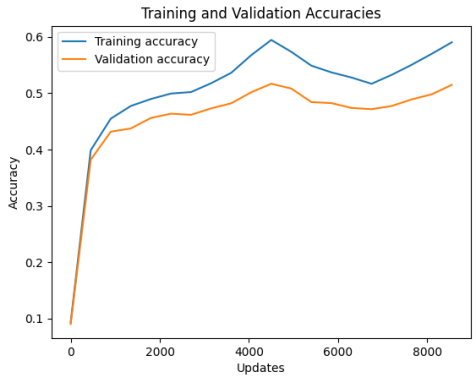


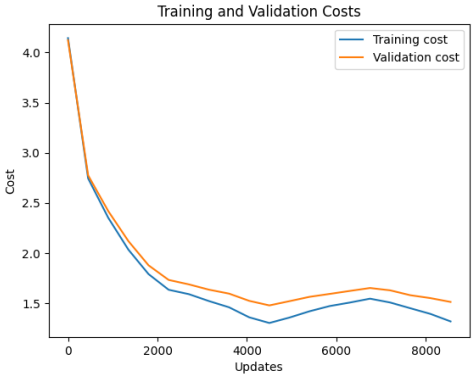
With Batch-Normalization

3-Layer Network

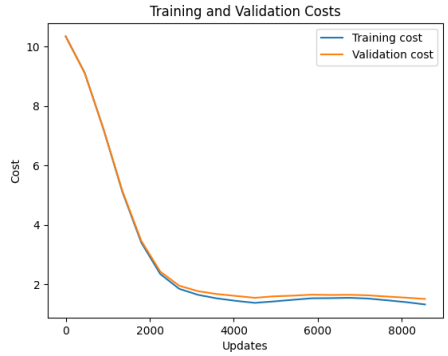
Settings: Lambda = 0.005, eta = [0.00001, 0.1], layers = [50, 50, 10], batch_size = 100, initialization: “He”.

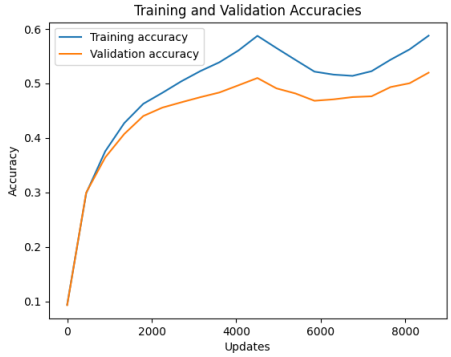
Final test accuracy: 0.534



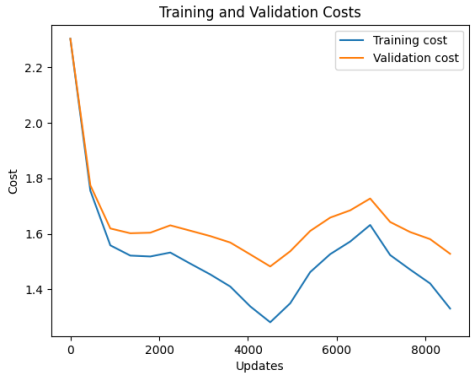


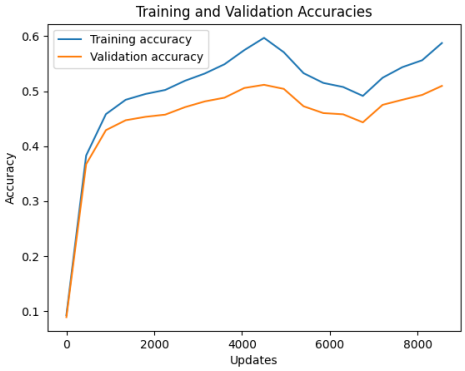
Initialization: sigma = 0.1, Final Test accuracy: 0.534



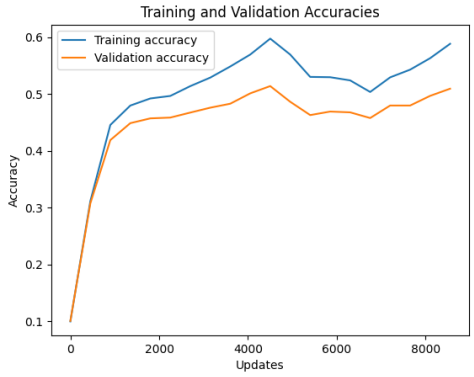


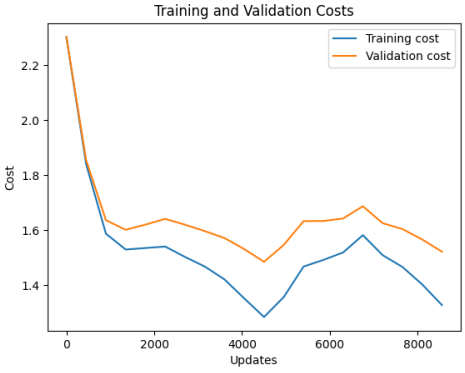
Initialization: sigma = 0.001, Final Test accuracy: 0.531





Initialization: sigma = 0.0001, Final Test accuracy: 0.527



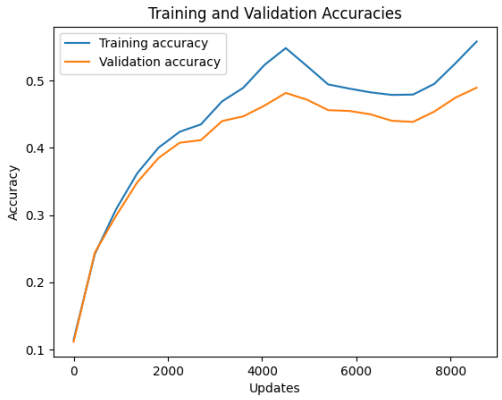


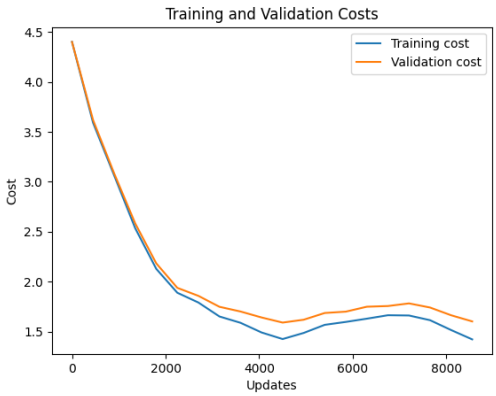
9-Layer Network

The following plots show the results for the following settings:

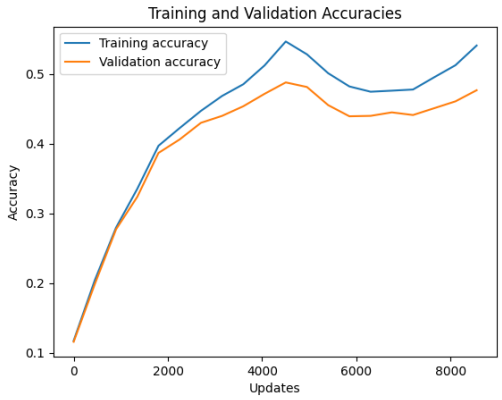
Lambda = 0.005, eta = [0.00001, 0.1], layers = [50, 30, 20, 20, 10, 10, 10, 10], batch_size = 100, initialization: “he”.

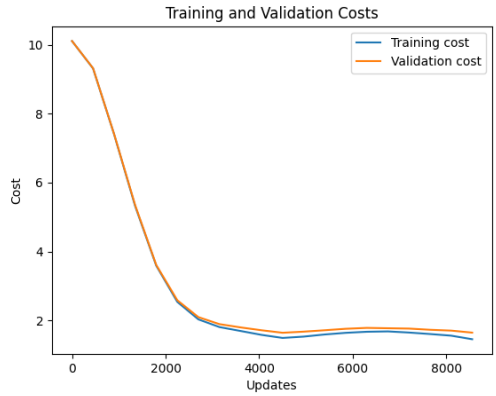
Final test accuracy: 0.523



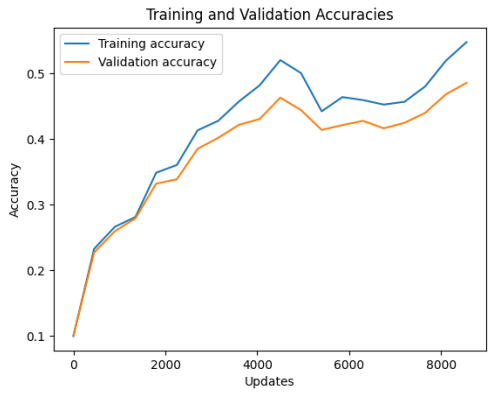


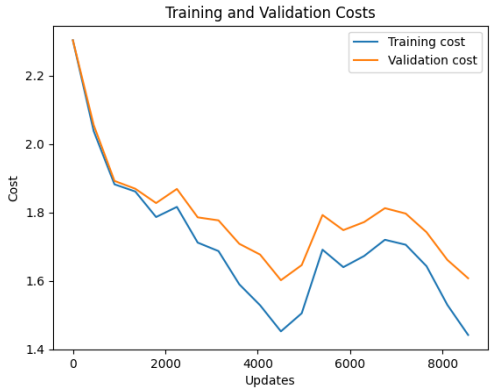
Initialization: sigma = 0.1, Final Test accuracy: 0.511



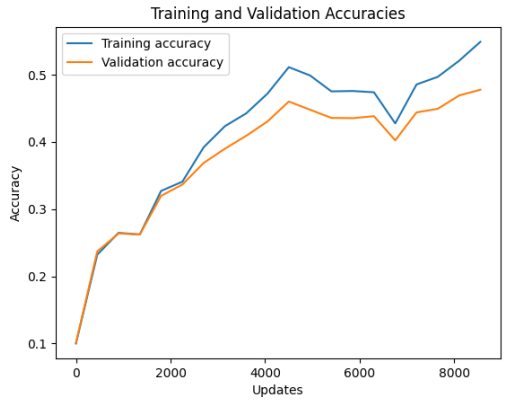


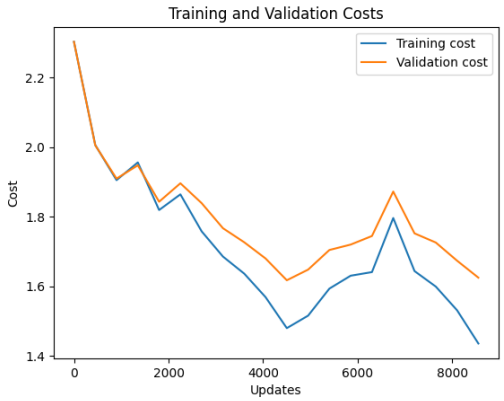
Initialization: sigma = 0.001, Final Test accuracy: 0.507





Initialization: sigma = 0.0001, Final Test accuracy: 0.509





Sensitivity to initialization

The batch normalization shows that for bigger networks the model is more robust and doesn't depend too much on the starting conditions. Without batch normalization, it depends a lot on the starting conditions, which one see's above with the different initializations with sigma and 'he'.