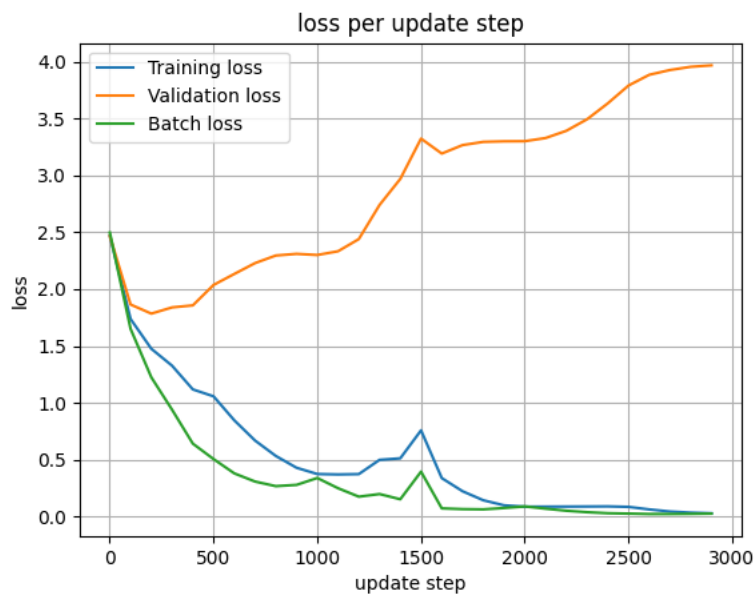


# Assignment 2

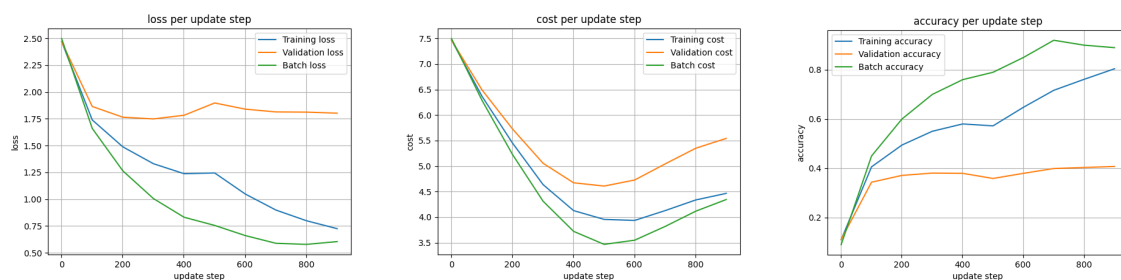
## Gradient computation check:

After having compared the gradient values with the `computationGradsNumSlow` method and my method, the gradients showed an extremely low difference. This made me believe the gradient calculation is correct.



This training was performed with the following settings:  $\lambda = 0.0$ ,  $n_{\text{batch}} = 100$ ,  $\eta_{\text{min}} = 0.00001$ ,  $\eta_{\text{max}} = 0.1$ ,  $\text{cycles} = 3$ ,  $\eta_s = 50000/n_{\text{batch}}$ . It clearly shows that the loss can get close to zero when no  $\lambda$  is used to compensate and reduce the extreme weights.

## Loss/Cost with cyclical learning rates and default values



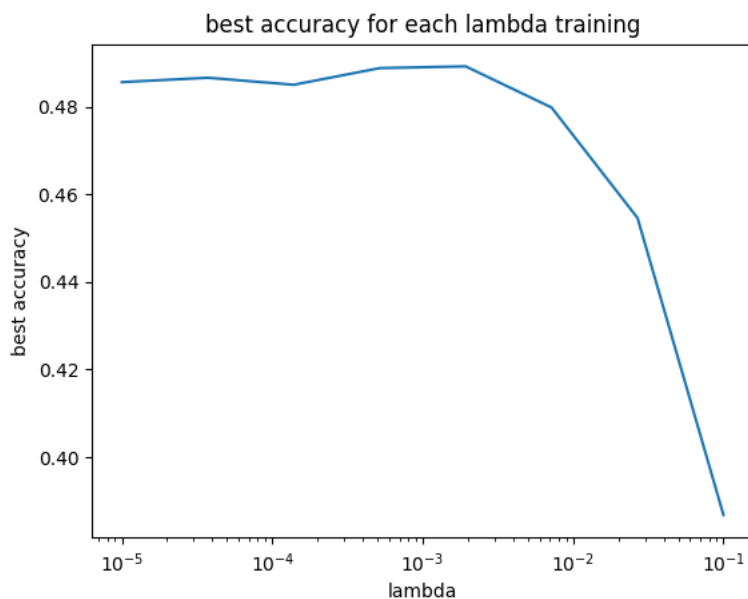
$\eta_{\text{min}} = 0.00001$ ,  $\lambda = 0.01$ ,  $\eta_{\text{max}} = 0.1$ ,  $\text{cycles} = 1$ ,  $\eta_s = 500$ ,  $n_{\text{batch}} = 100$   
plotted every 100<sup>th</sup> value  
accuracy\_train: 0.838, accuracy\_validation: 0.4048, accuracy\_batch: 0.79



lambda\_ = 0.01, n\_batch = 100, eta\_min = 0.00001, eta\_max = 0.1, cycles = 3, eta\_s = 800  
 plotted every 100<sup>th</sup> value  
 accuracy\_train: 0.9682, accuracy\_validation: 0.397, accuracy\_batch: 0.97

### Coarse lambda search

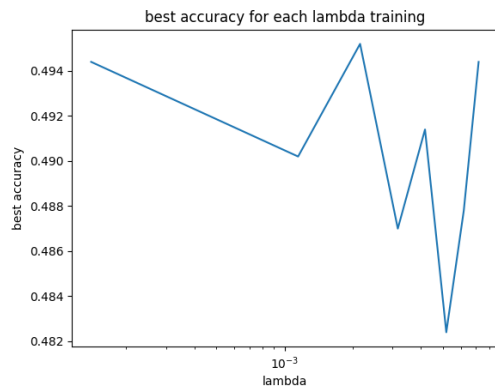
The optimal lambda was searched in the range from 0.0001 and 0.1 with 8 steps in between. The following plot shows the best accuracy on the validation set for each lambda used during training. The following settings were used: n\_batch: 100, eta\_min = 0.00001, eta\_max = 0.1, cycles = 1, train set = 45000, validation set = 5000, eta\_s = trainset/n\_batch



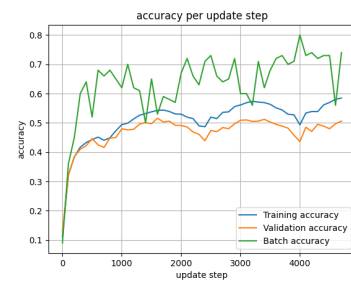
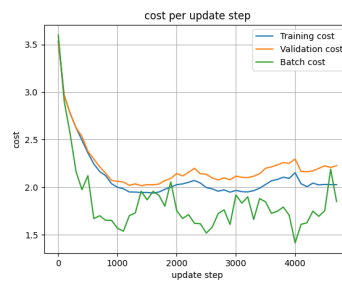
The three best accuracies were reached with the following lambdas:  $3.72759372e-05$ ,  $1.93069773e-03$  and  $5.17947468e-04$

### Fine lambda search

During the fine search for the optimal lambda following parameters were used: n\_batch = 100, eta\_min = 0.00001, eta\_max = 0.1, cycles = 1, eta\_s = 500, lambda\_min =  $1.4e-04$ , lambda\_max =  $7.2e-03$ , steps between lambda min and max = 8,  
 Best lambdas:  $2.15714286e-03$ ,  $7.20000000e-03$  and  $1.40000000e-04$



## Best configuration



lambda\_ = 2.15714286e-03, n\_batch = 100, eta\_min = 0.00001, eta\_max = 0.1, cycles = 3, eta\_s = 800,

Used data: 49000 train set, 1000 validation set, 10000 test set

accuracy train: 0.5942857142857143, accuracy validation: 0.508, accuracy test: 0.5021