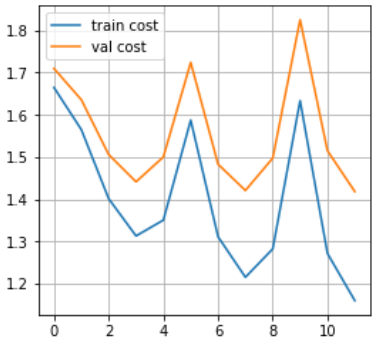
# Assignement 2: Report bonus

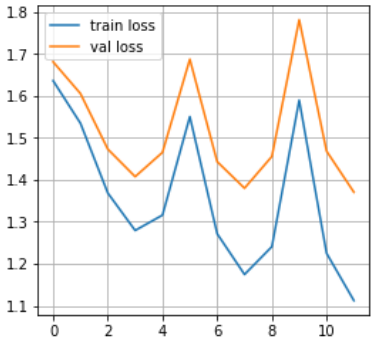
## a) Improvements

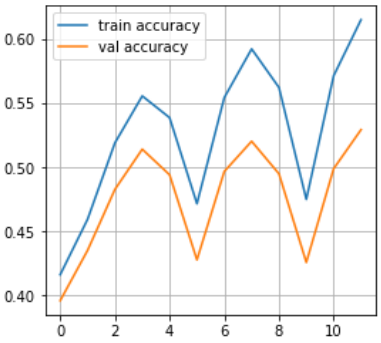
| | lambda | batch | Hidden nodes | eta_min | eta_max | n_s | acc_train | acc_valid | acc_test |
|---|---|---|---|---|---|---|---|---|---|
| **Basic** | 0.00047 | 100 | 50 | 1e-5 | 0.1 | 900 | 0.6692 | 0.5188 | **0.5117** |
| **1) More hidden nodes** | 0.01 | 100 | 100 | 1e-5 | 0.1 | 900 | 0.615289 | 0.5294 | **0.5186** |
| **2) Data augmentation** | 0.00047 | 100 | 50 | 1e-5 | 0.1 | 900 | 0.61055 | 0.5274 | **0.5157** |
| **3) Ensemble of 5 NNs from local minimas** | 0.00047 | 80 | 50 | 1e-5 | 0.1 | 900 | 0.6706444 | 0.5318 | **0.5126** |
| **4) All combined** | 0.01 | 100 | 100 | 1e-5 | 0.1 | 900 | 0.577533 | 0.54 | **0.5357** |

In addition to these I tried Boosting on 5 NNs where each was trained on 10000 samples but it didn't give any improvements.

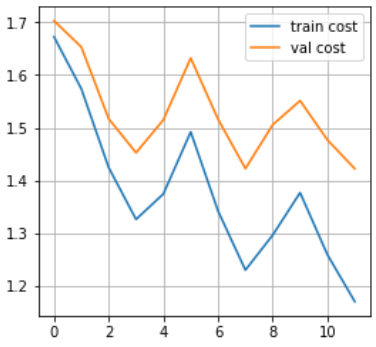
### 1) More hidden nodes

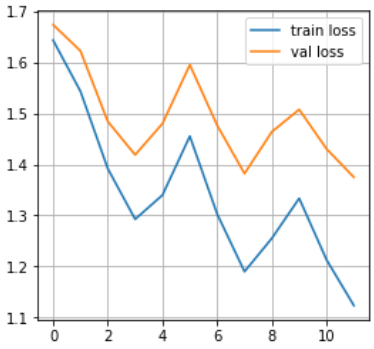


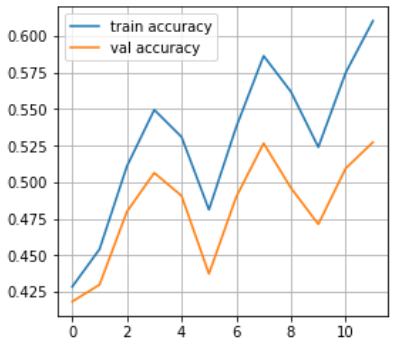



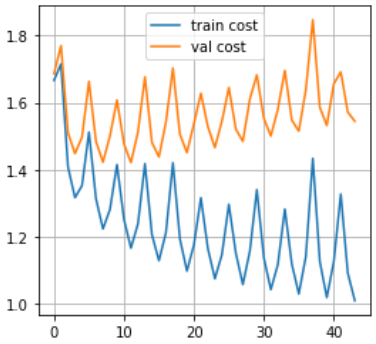
### 2) Data augmentation

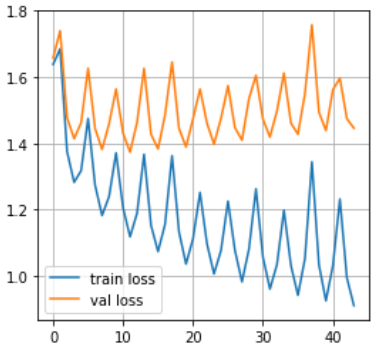


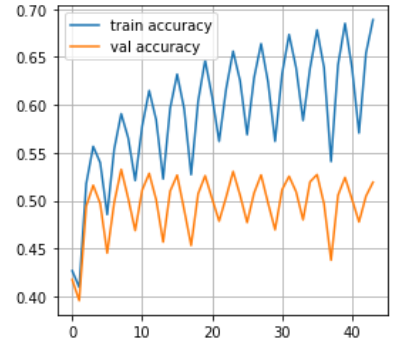



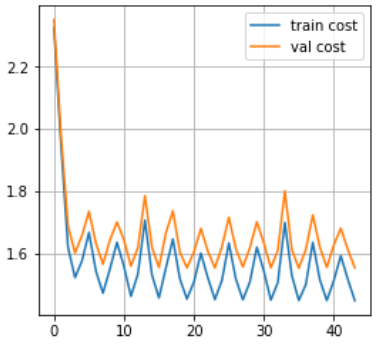

### 3) Ensemble of 5 NNs from local minimas

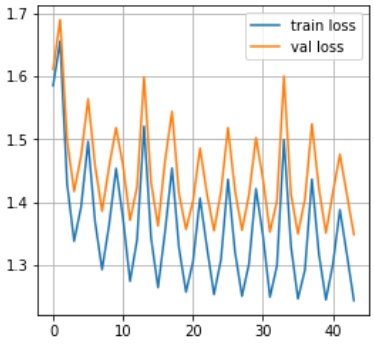


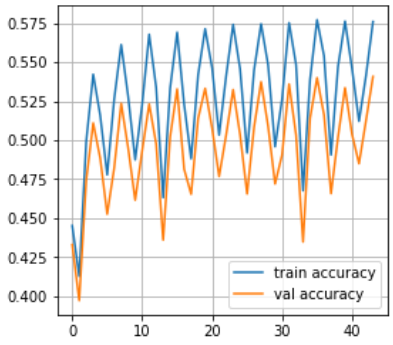




### 4) All improvements combined




From the 3 improvements presented in the table above adding more hidden nodes and increasing regularization happened to deliver the best improvements. By increasing the number of hidden nodes, theoretically, any nonlinear function can be approximated with the NN. In our case the network with 100 hidden nodes captures the nonlinearities of the data better than the one with only 50 hidden nodes.

### b) Best parameters for the cyclical learning rate( LR Test)

In this part we follow the suggestions presented in [Smith, 2015] to find the best parameter for the cyclical learning rate.

According to the author the stepsize is normally chosen as *n_s = n / n_batch *k* where n is the number of training samples and k can be a number 2-10.
In order to choose the best range for the cyclical learning rate one does a so called LR Test where the learning rate is increased linearly during all update steps. Eta_min is choosed to be the lowest learnign rate where the network converges and eta_max the first learning rate where the accuracy on the validation suffers.

In our case eta_min = 1e-5 and eta_max = 0.072 which are very close with the values suggested in the assignement.