

Net1	Net2	Net3
Train error: 0.4357	Train error: 0.4428	Train error: 0.4058
Validation error: 0.5167	Validation error: 0.5225	Validation error: 0.4960
Test set error: 0.5122	Test set error: 0.5217	Test set: 0.4952
epochs: 195	epochs: 165	epochs: 323

Table 1: Classification errors for three different networks. All networks have the same *early stopping* criteria. All networks are also updating by using *log loss* instead of mean-error squared. Network 1 and 3 has the same layer layout but network 3 also has added L2 regularization. Network 2 has addition hidden layer than the other two networks.

Theoretical: *Early stopping* is a method that helps networks from overfitting. The network will stop the training at an epoch when the validation performance starts performing worse than the training. *Regularization* is another technique for reducing overfitting by penalizing the parameters so they don't reach large enough values.

Experimental results: The investigation consisted of three different types of networks. Table 1 shows each of the networks reached an early stopping at different epochs. Network 2 stopped training at earlier epoch because it is a deeper network and hence more prone to overfit causing bad performance on the validation set. Network 3 stopped training at a much later epoch because of regularization, hence allowing the network to train at longer epochs reaching better generalization results.