ANN Lab 1

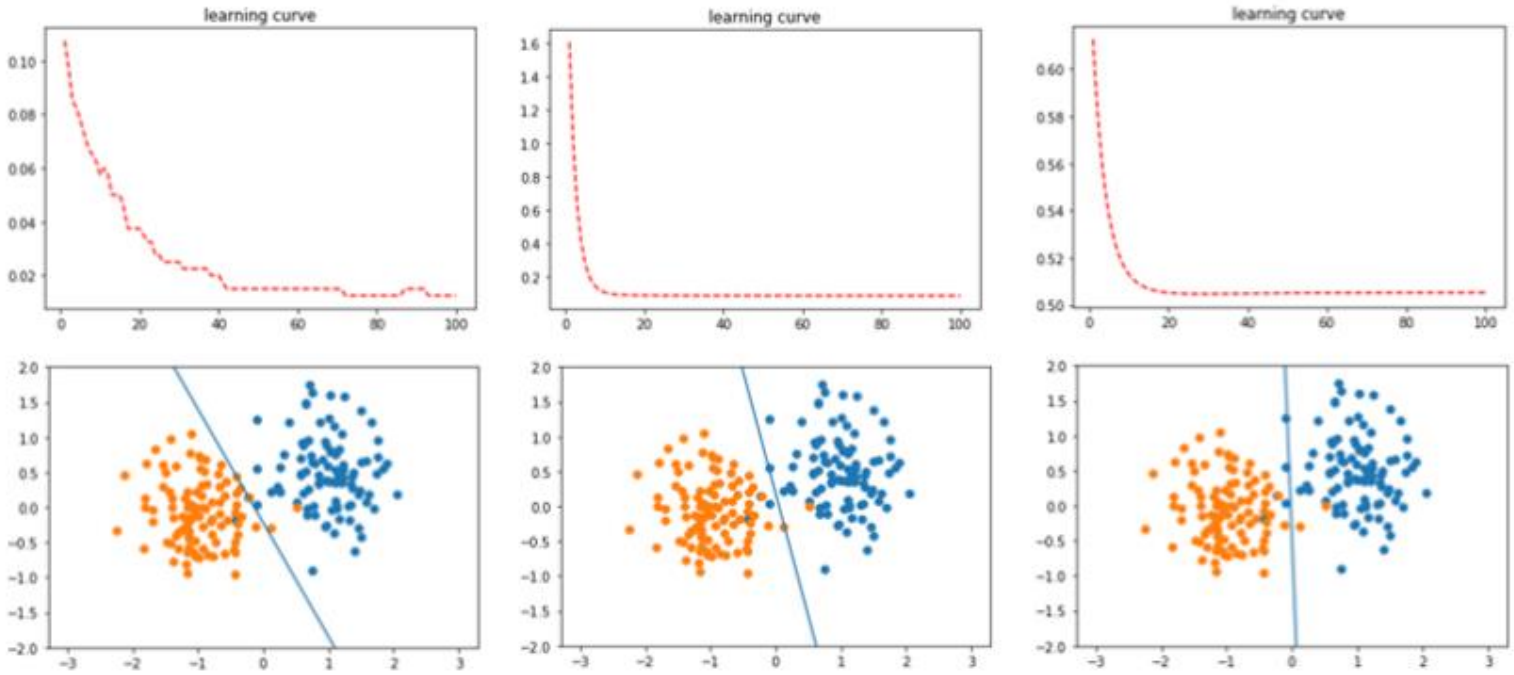
Forward Learning and Backward Propagation, Perceptron Learning and Delta Rule, One-Layer, Two-Layer and Multi-Layer Learning for classification and regression, Encoder Problem

Ravi Bir|ravib@kth.se
Bharat Sharma|bsharma@kth.se
Qiao Ren|qiaor@kth.se

February 5, 2020

Single-layer perceptron

- Delta Batch converged fastest, then delta sequential, then perceptron
- Learning rate of 0.001 gave the fastest convergence for all 3



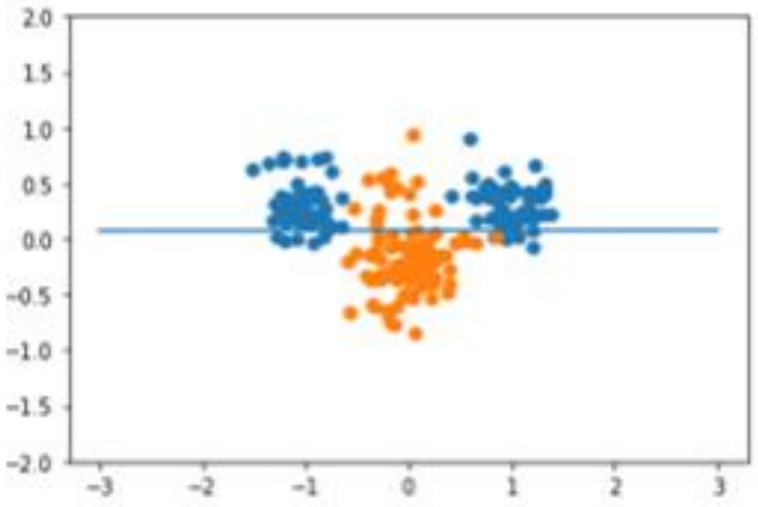
Figur 1: Learning curve Perceptron

Figur 2: Learning curve Delta Rule Batch

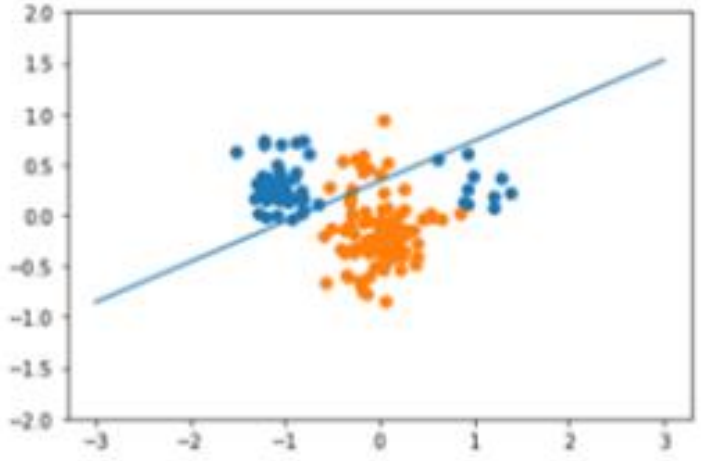
Figur 3: Learning curve Delta Rule Sequential

Non Linearly Separable Data

- Obvious that linear boundary not sufficient- need 2 layer perceptron



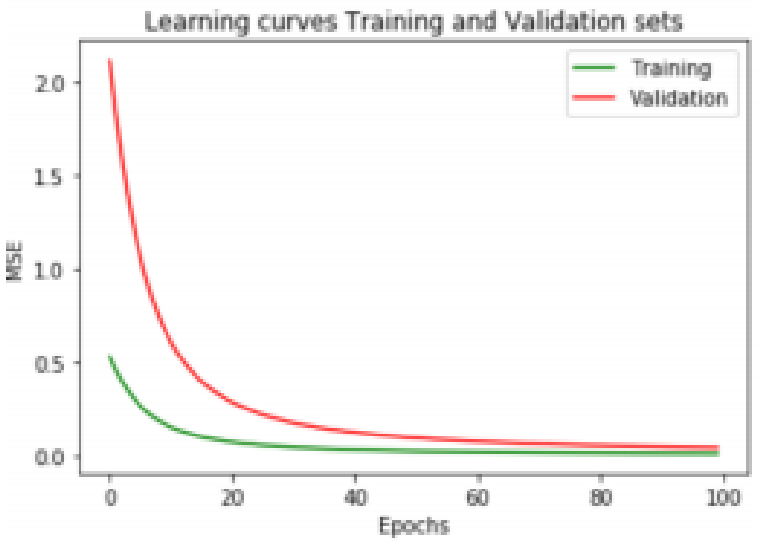
Figur 4: Nothing Removed



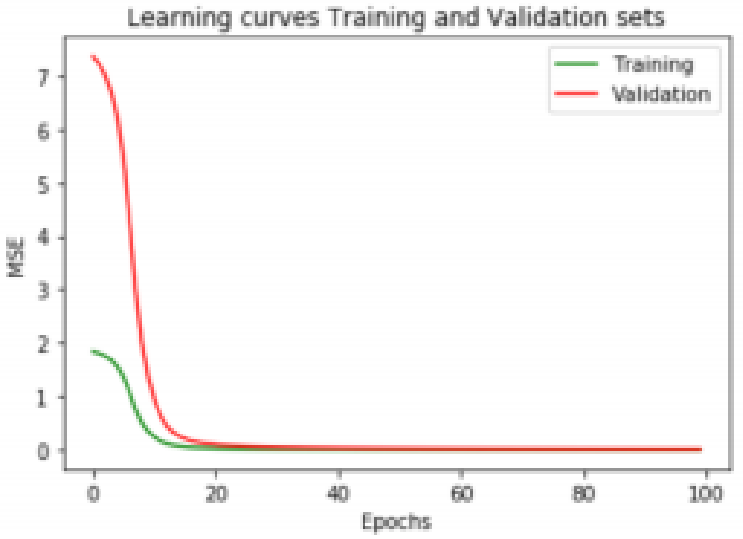
Figur 8: Remove subsets of class A

Classification with a two-layer perceptron

- More nodes means faster convergence, but chance of overfitting
- Learning rate 0.009



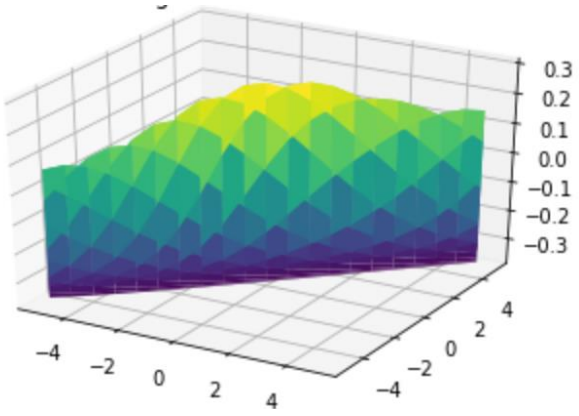
Figur 9: Learning curves, 4 nodes



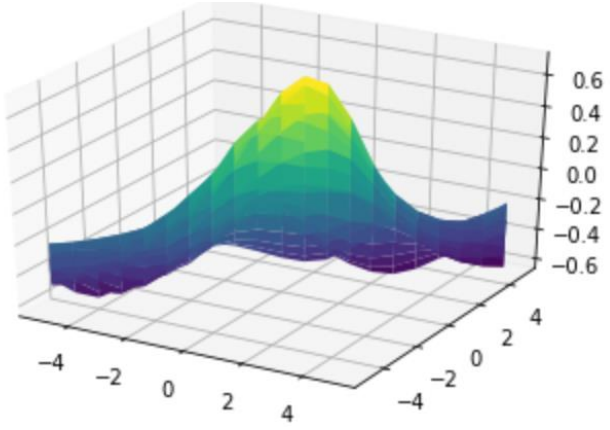
Figur 10: Learning curves, 25 nodes

Encoder and Function approximation

- Encoder used for dimensionality reduction and reducing the number of hidden nodes cause the encoder to take more epochs to converge.
- For learning rate 0.009, converges in 20-30 epochs.
- Each input used for activation of hidden layer.
- Function approximation -



2 Hidden nodes



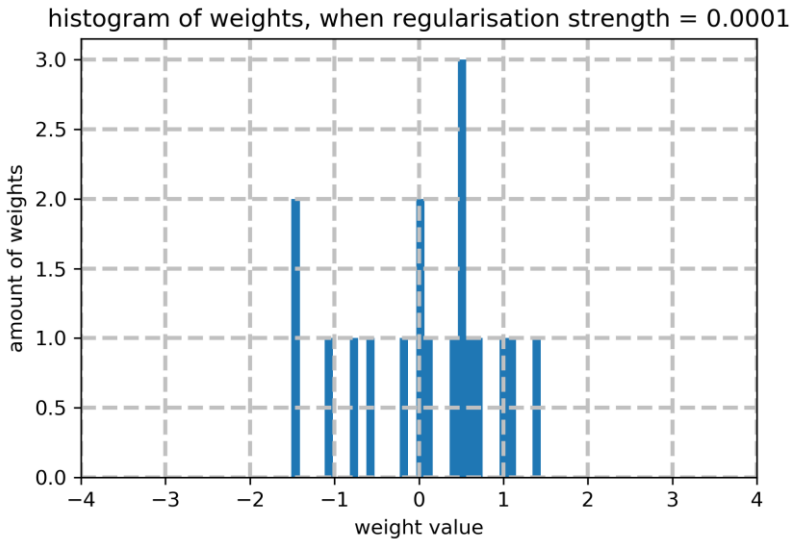
25 hidden nodes

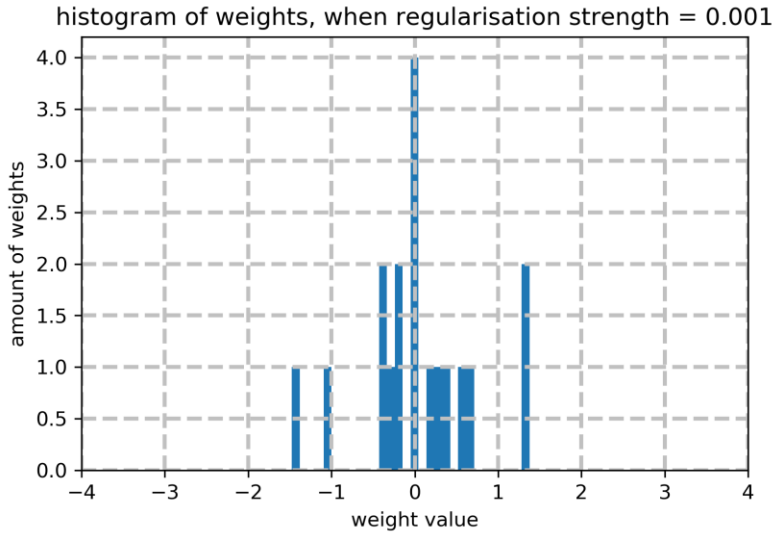
Amount of Training data(%)	Error(MSE)
80	0.004
60	0.006
40	0.008
20	0.01

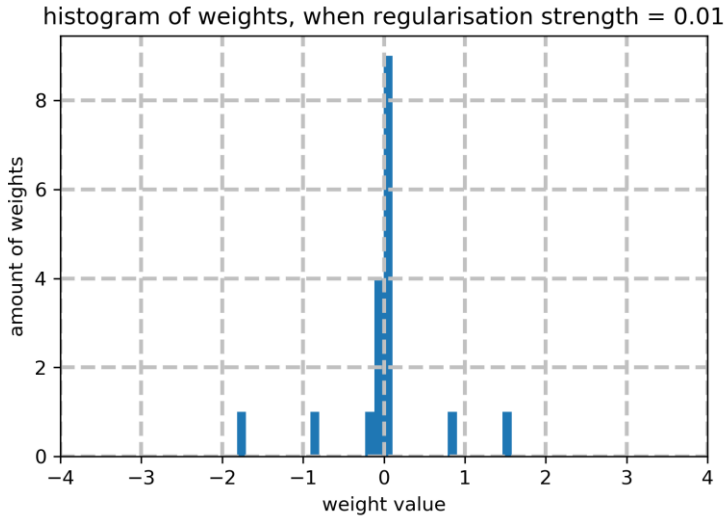
Tabell 3: Performance of 22 node model. η - 0.009

Two layer Model

- large regularization strength reduces lots of weights to small values (close to zero)

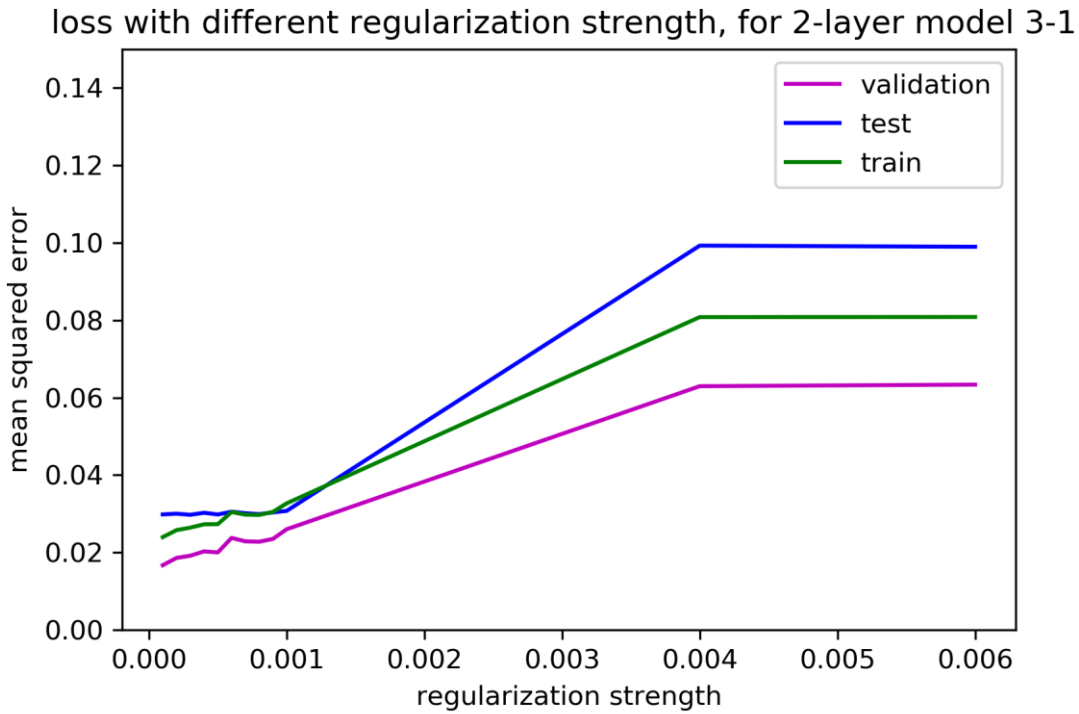


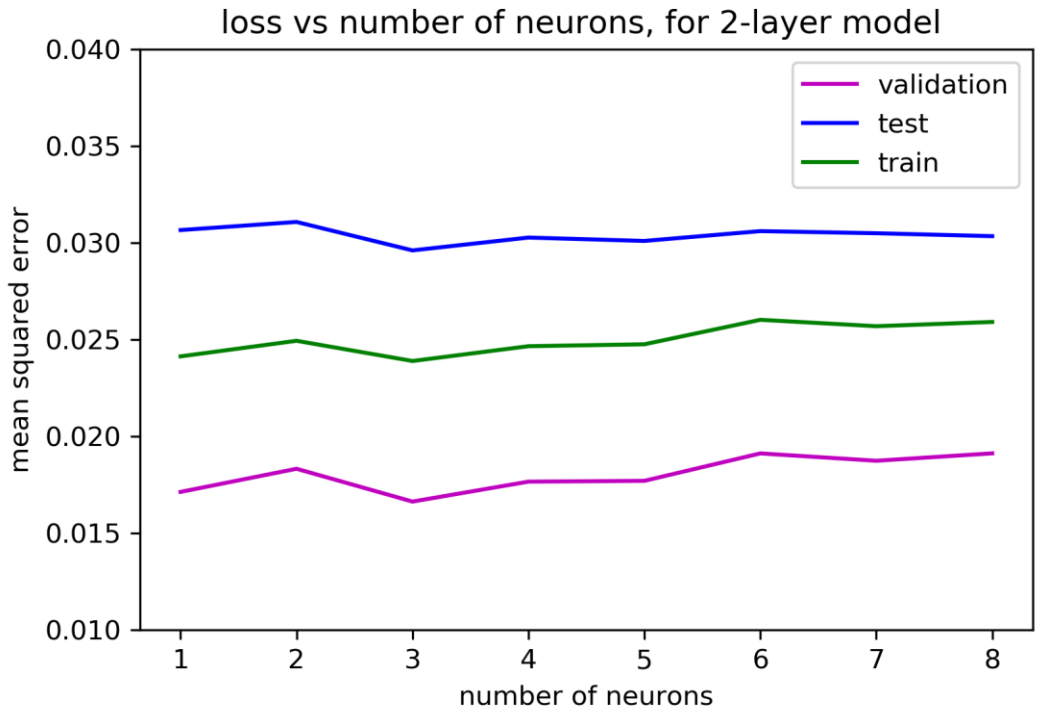




Two layer Model

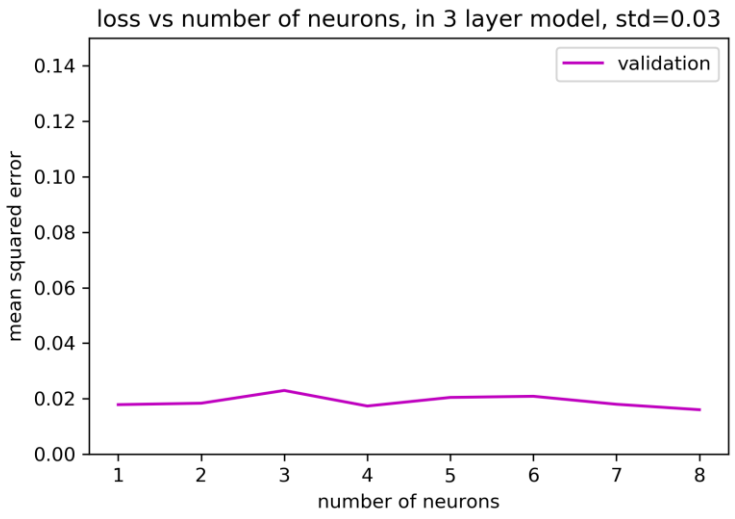
- regularization term penalize the weights, reduce the complexity
- we chose regularization strength= 0.0001 (this is a simple model)

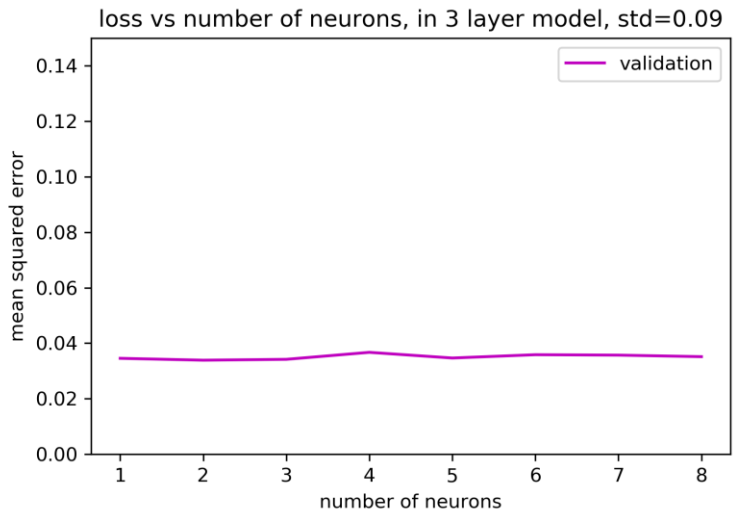


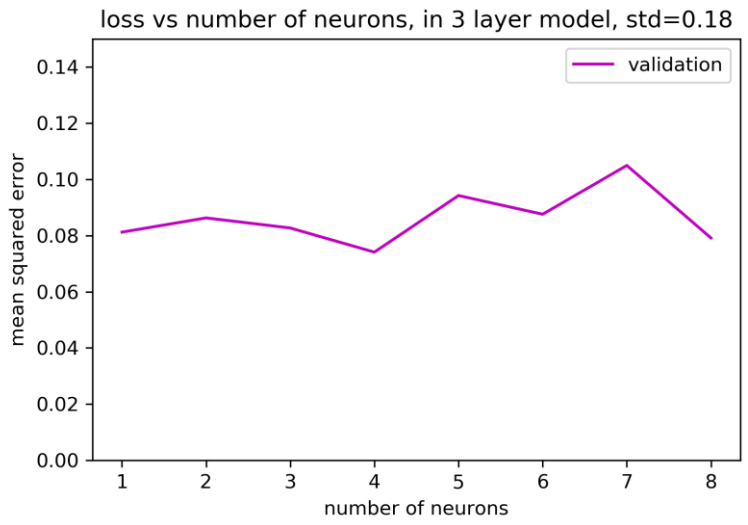


Three layer Model

- larger noise raises up the MSE on validation data
- the optimal amount of neurons on 2nd layer varies when the noise is different







Comparison on Two-Layer and Three-Layer Model

- each model has been run for 10 times. We recorded the average.
- MSE: 3-2-1 is lower computational time: 3-1 < 3-2-1
- could not generalize the conclusion, because only run 10 times

Model	Parameters	Validation MSE	Test MSE	Computational Time (second)
2 layer model: 3-1	lr=0.001, regu=0.0001	0.0305	0.0417	6.5209
3 layer model: 3-2-1	lr=0.001, regu=0.0001	0.0291	0.0393	6.8346