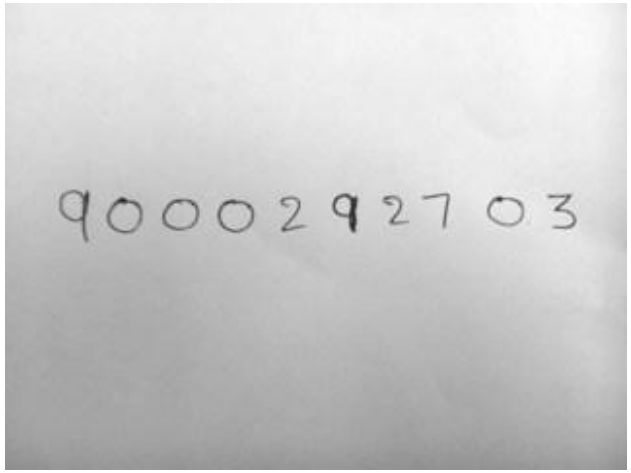
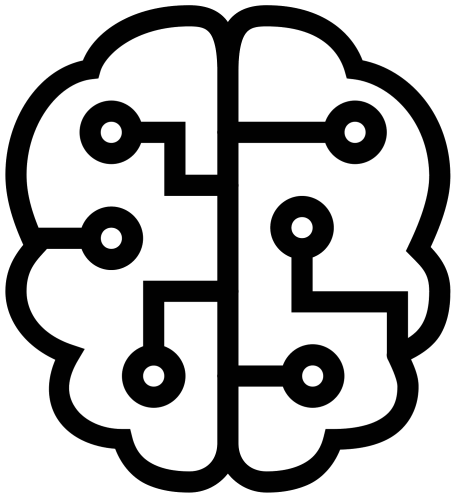
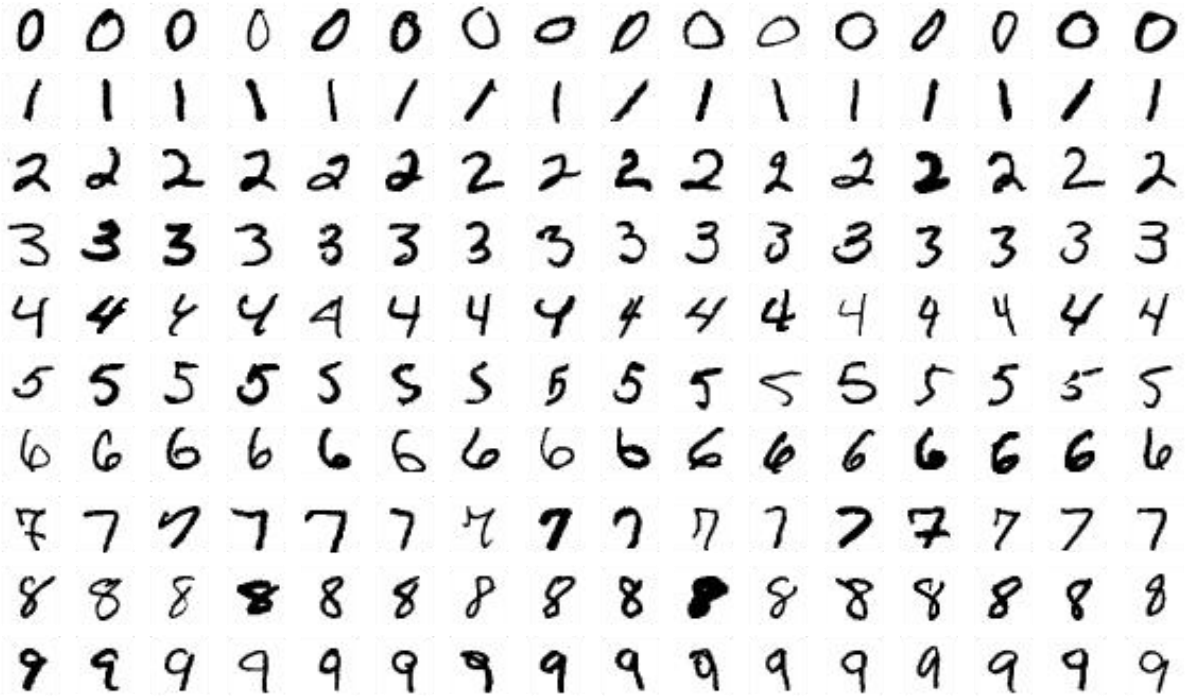
Studi Kasus

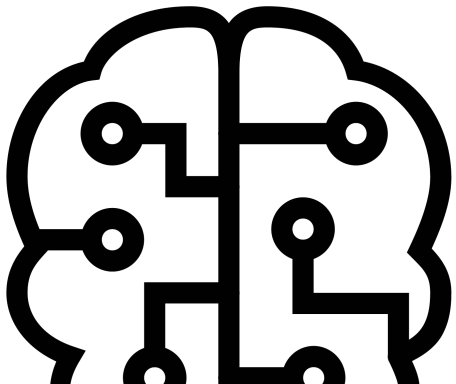




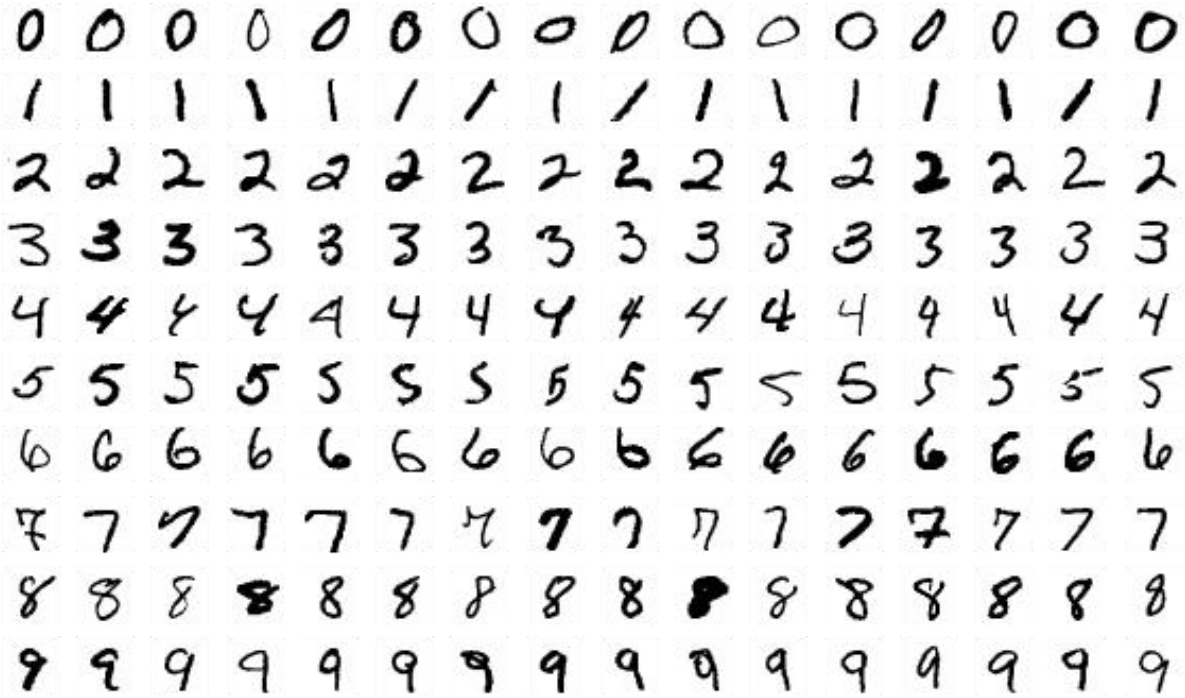
9000292703

Studi Kasus





Studi Kasus

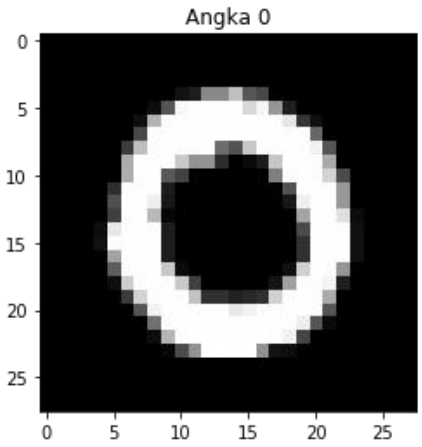


MNIST–Dataset

Ribuan Data
Training dan Test

Openly
Published

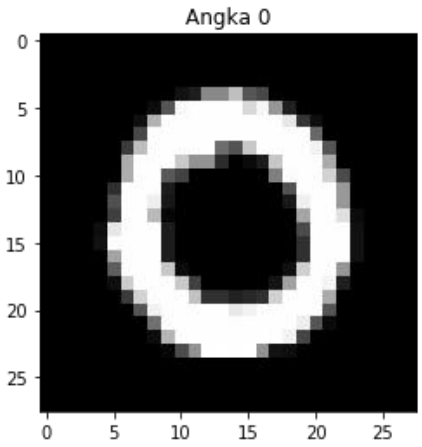
Label Encoding



Model
Deep
Network
Kita

0

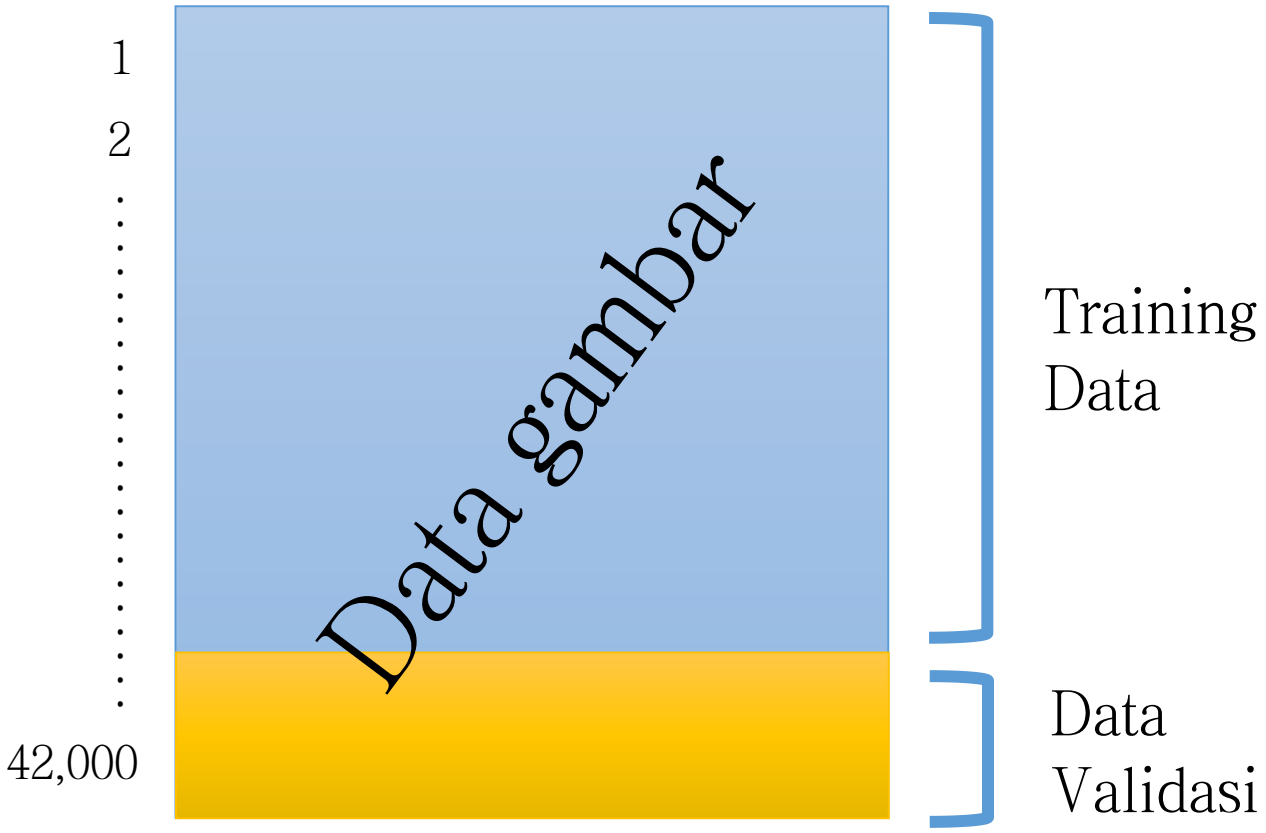
One Hot Encoding



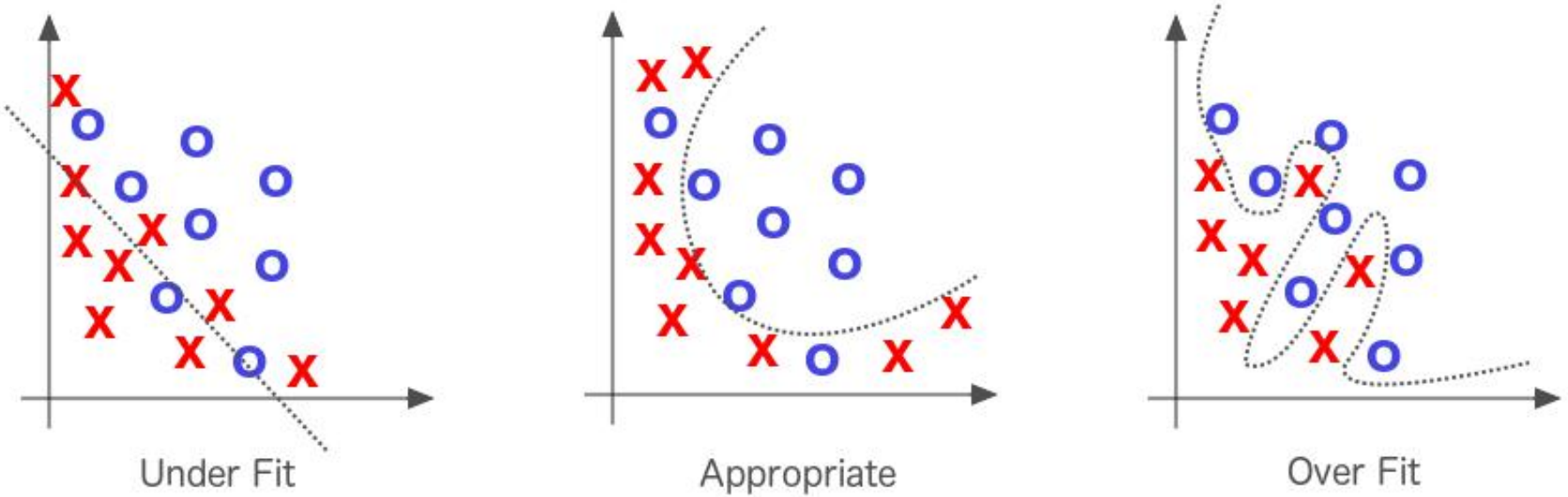
Model
Deep
Network
Kita

0	0.75
1	0.02
2	0.01
3	0.01
4	0.03
5	0.10
6	0.02
7	0.01
8	0.03
9	0.02

Data Training dan Data Validasi



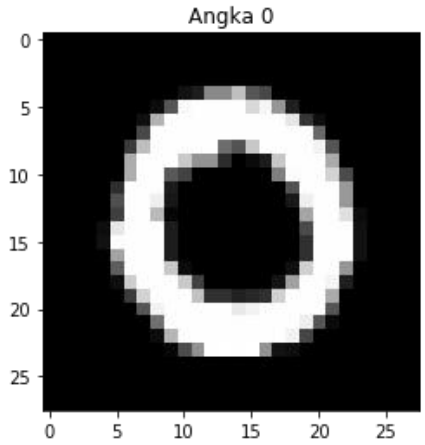
Underfitting dan Overfitting



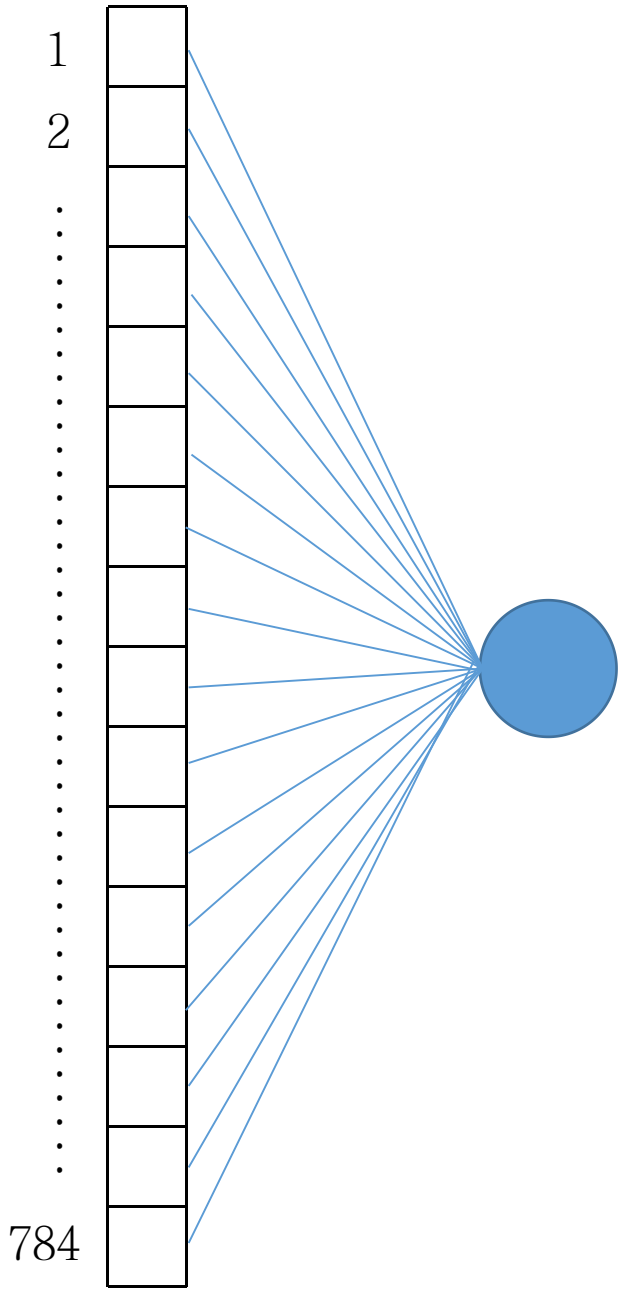
CNN dan ANN, bedanya apa?

- Punya neuron / bias yang dapat "belajar"
- Menerima input, melakukan operasi, keluar class
- Membutuhkan loss-function

Total Weight = 768

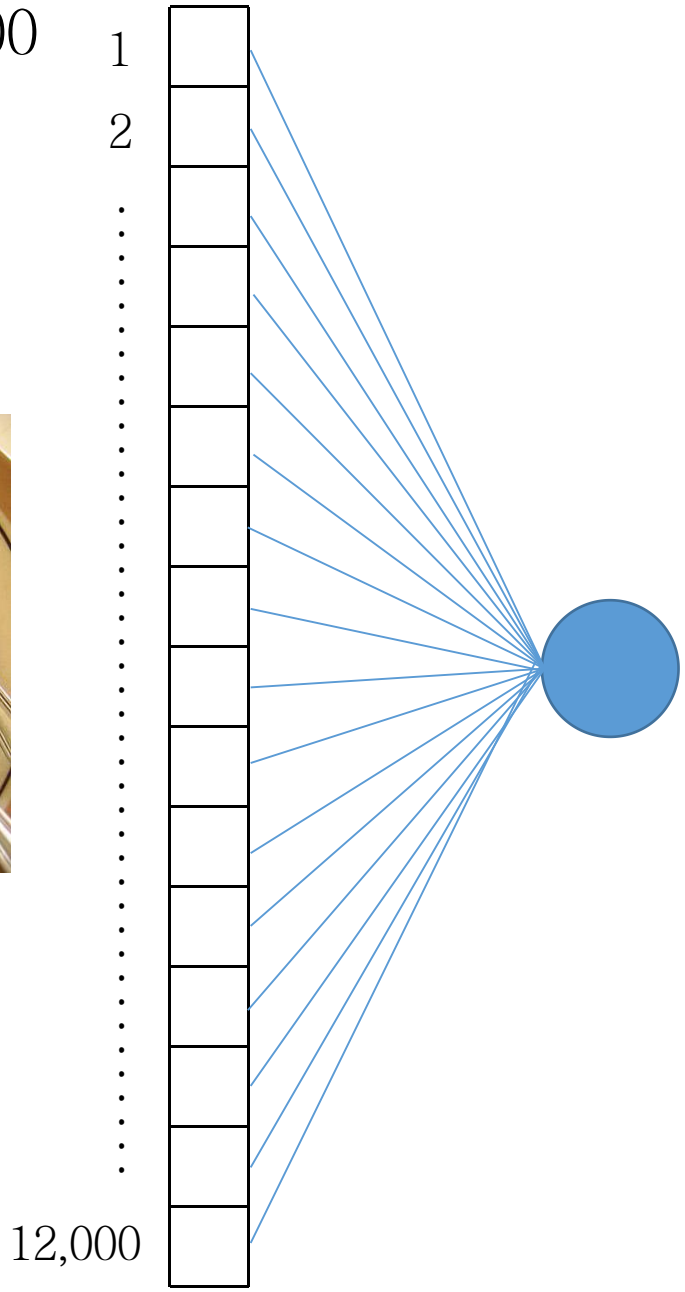


28x28



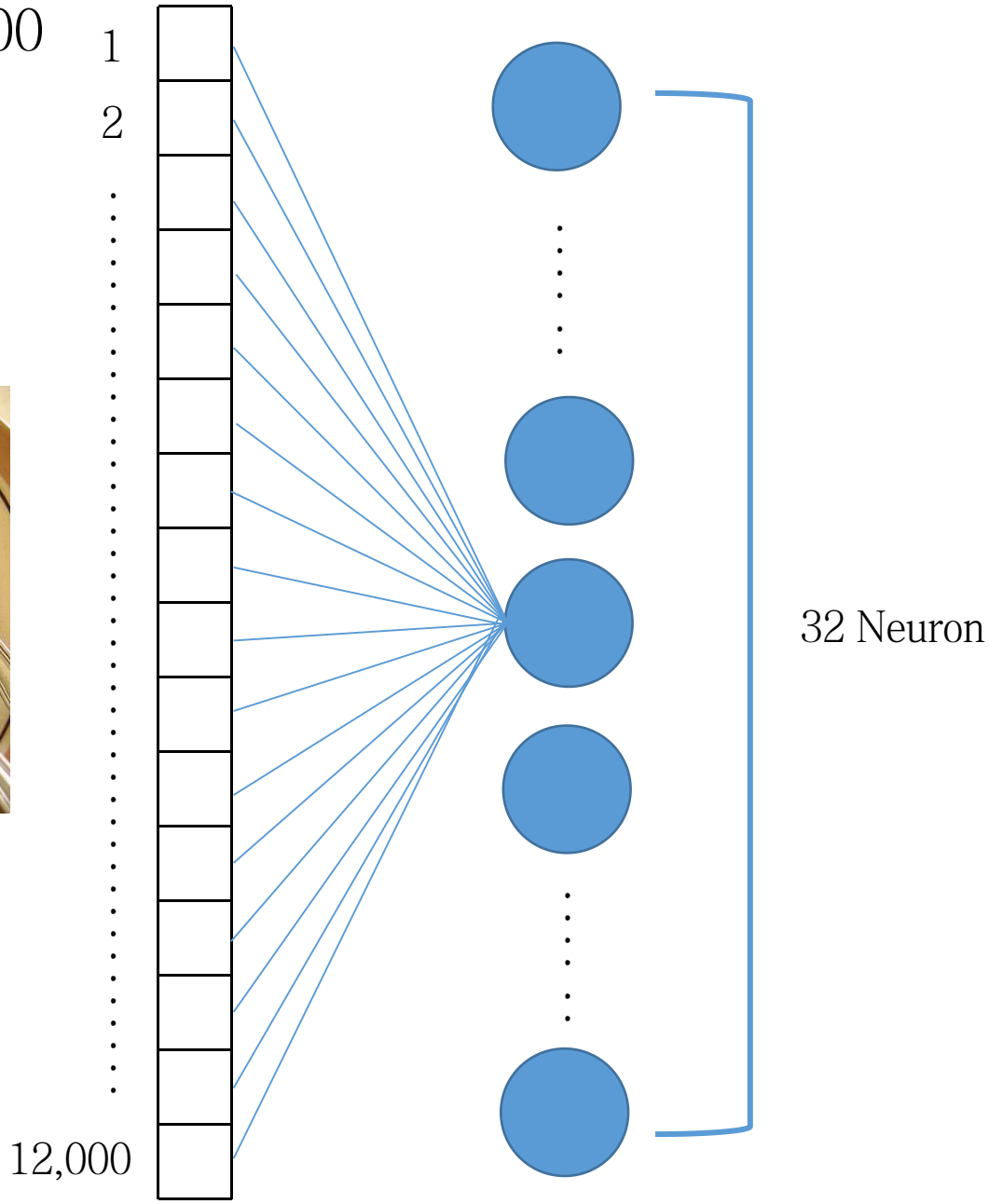
Total Weight = 12,000

200x200x3



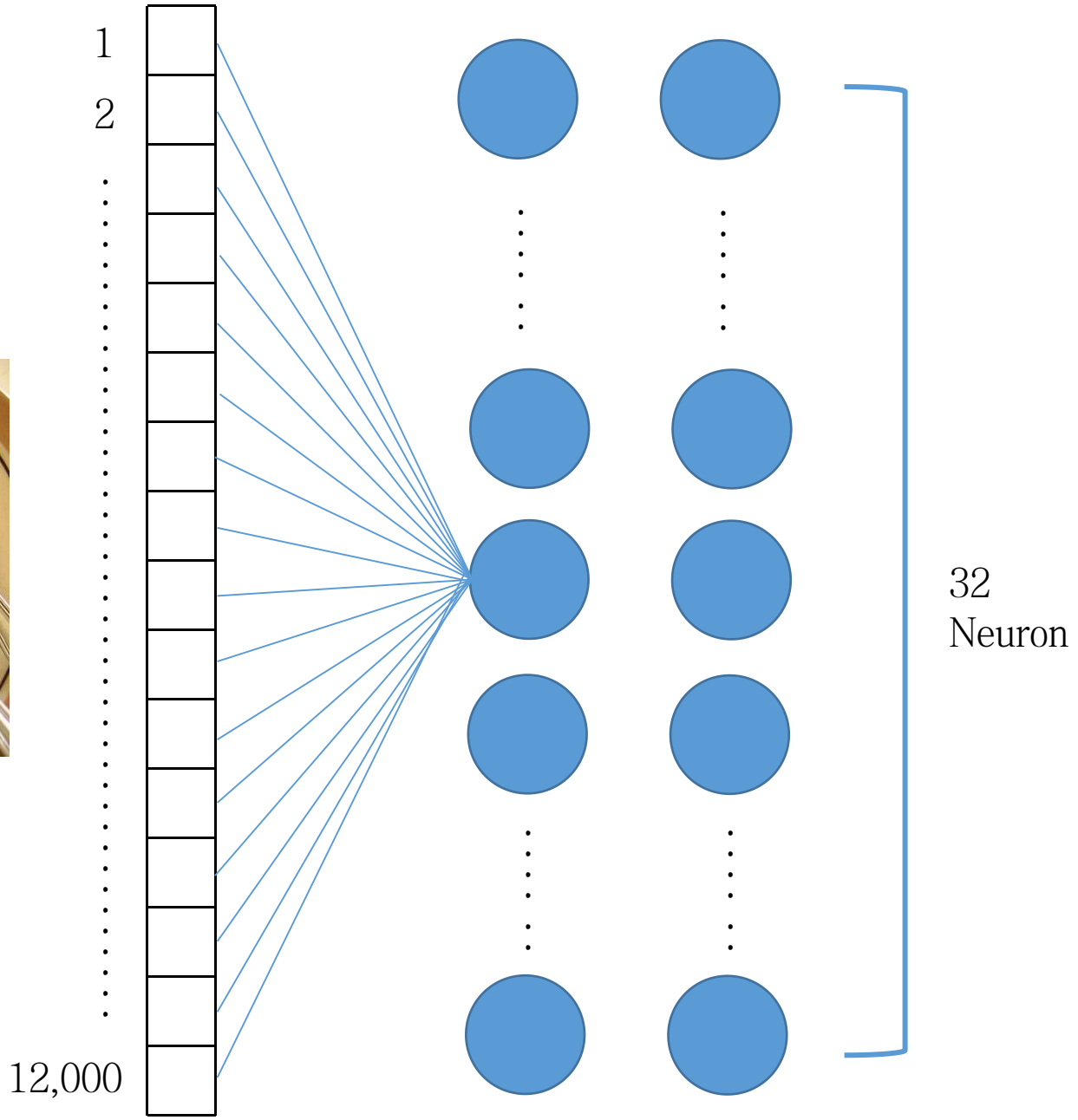
Total Weight = 384,000

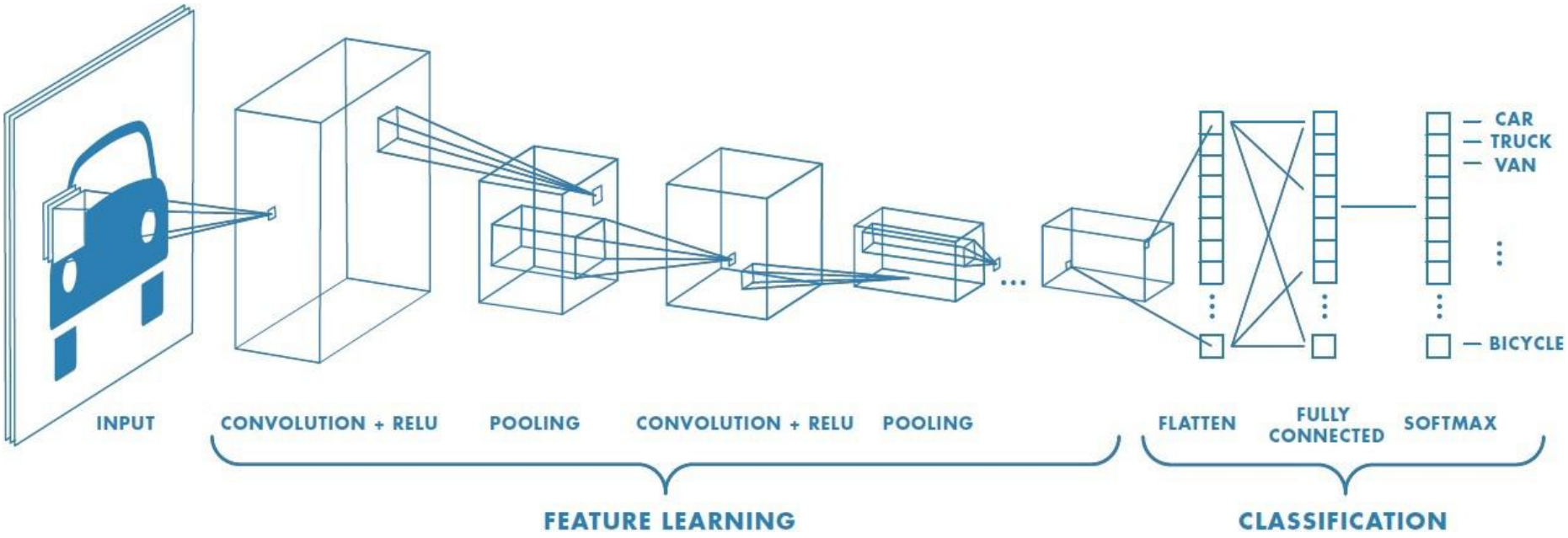
200x200x3

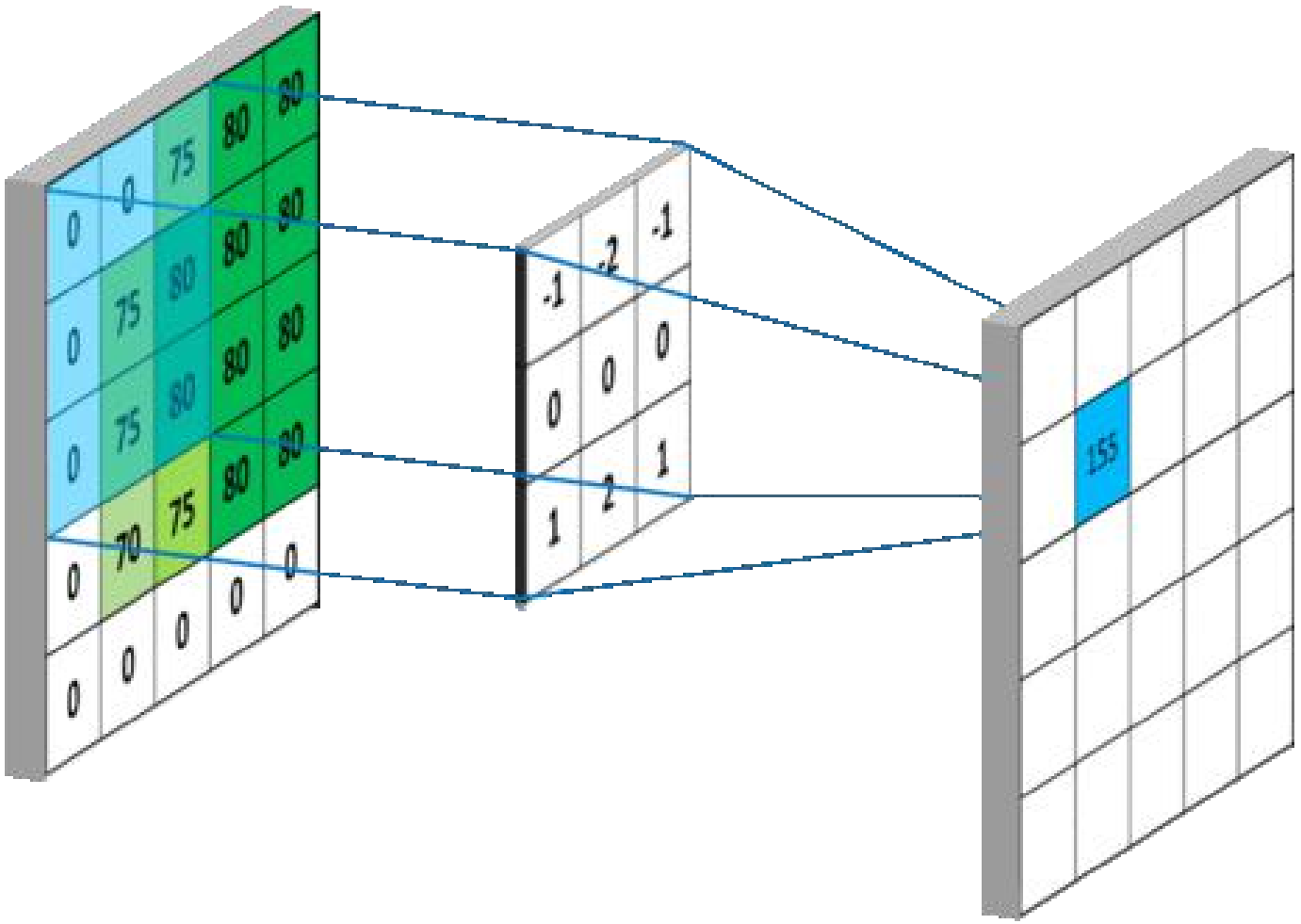


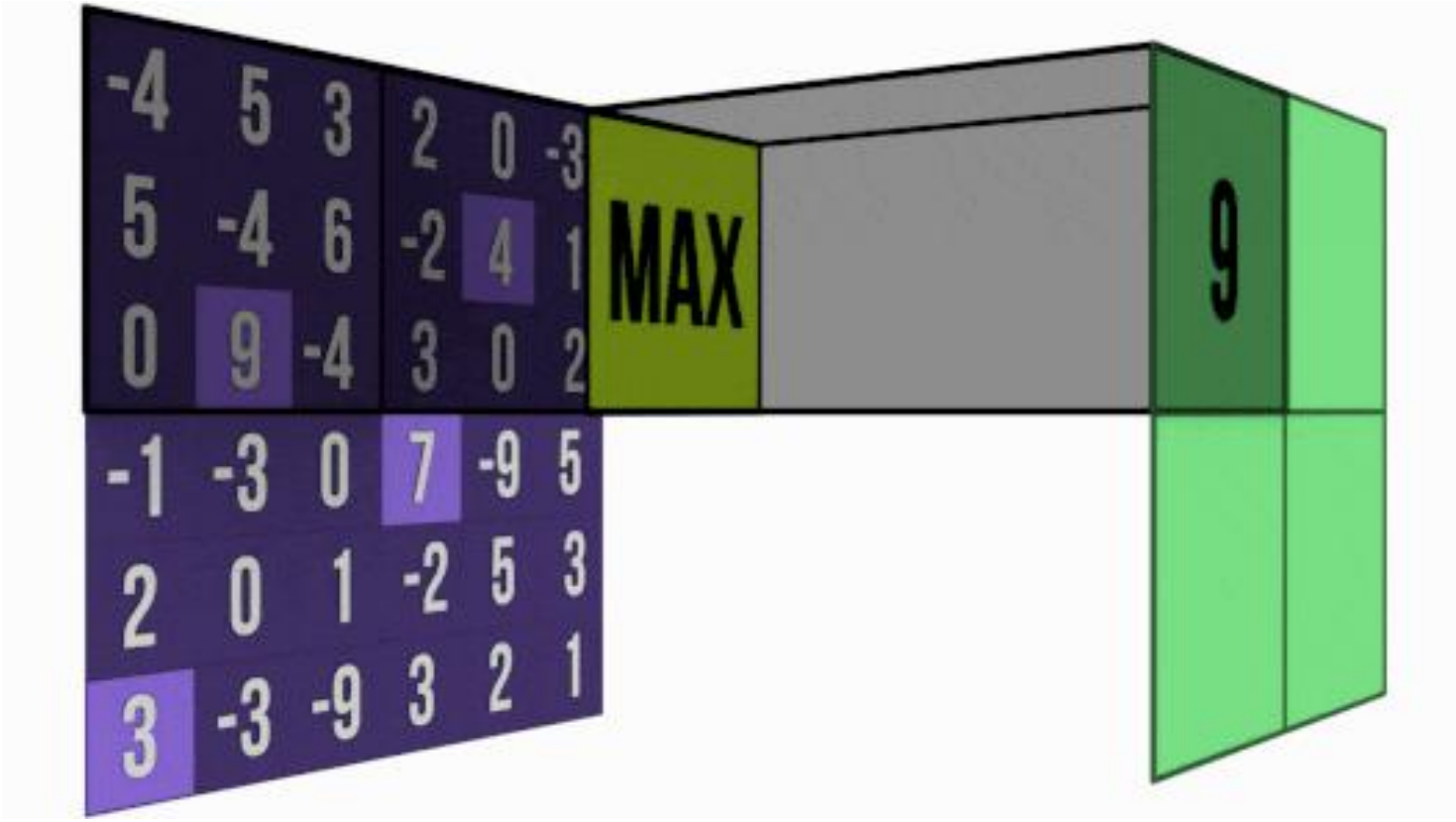
Total Weight =
12,228,000

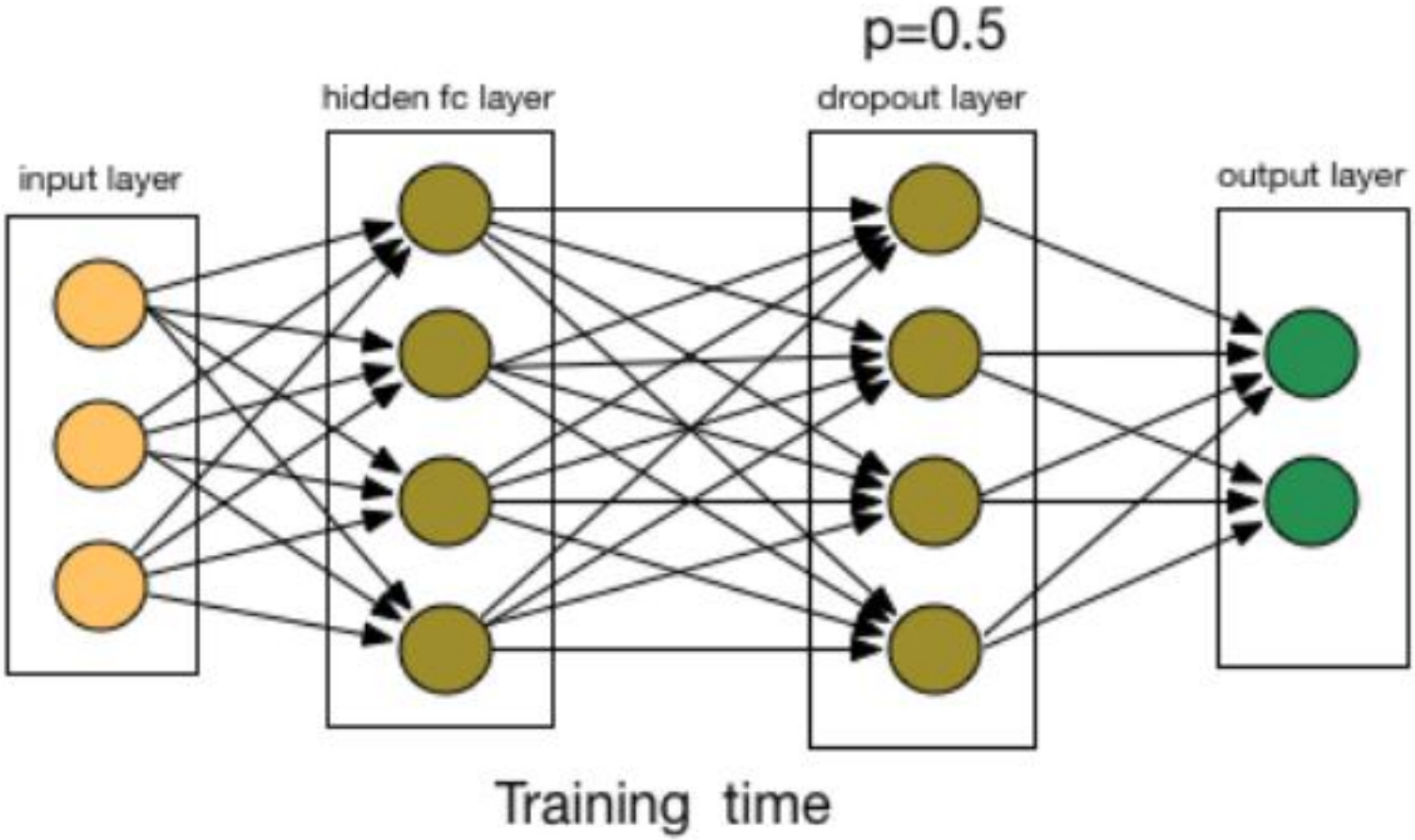
200x200x3

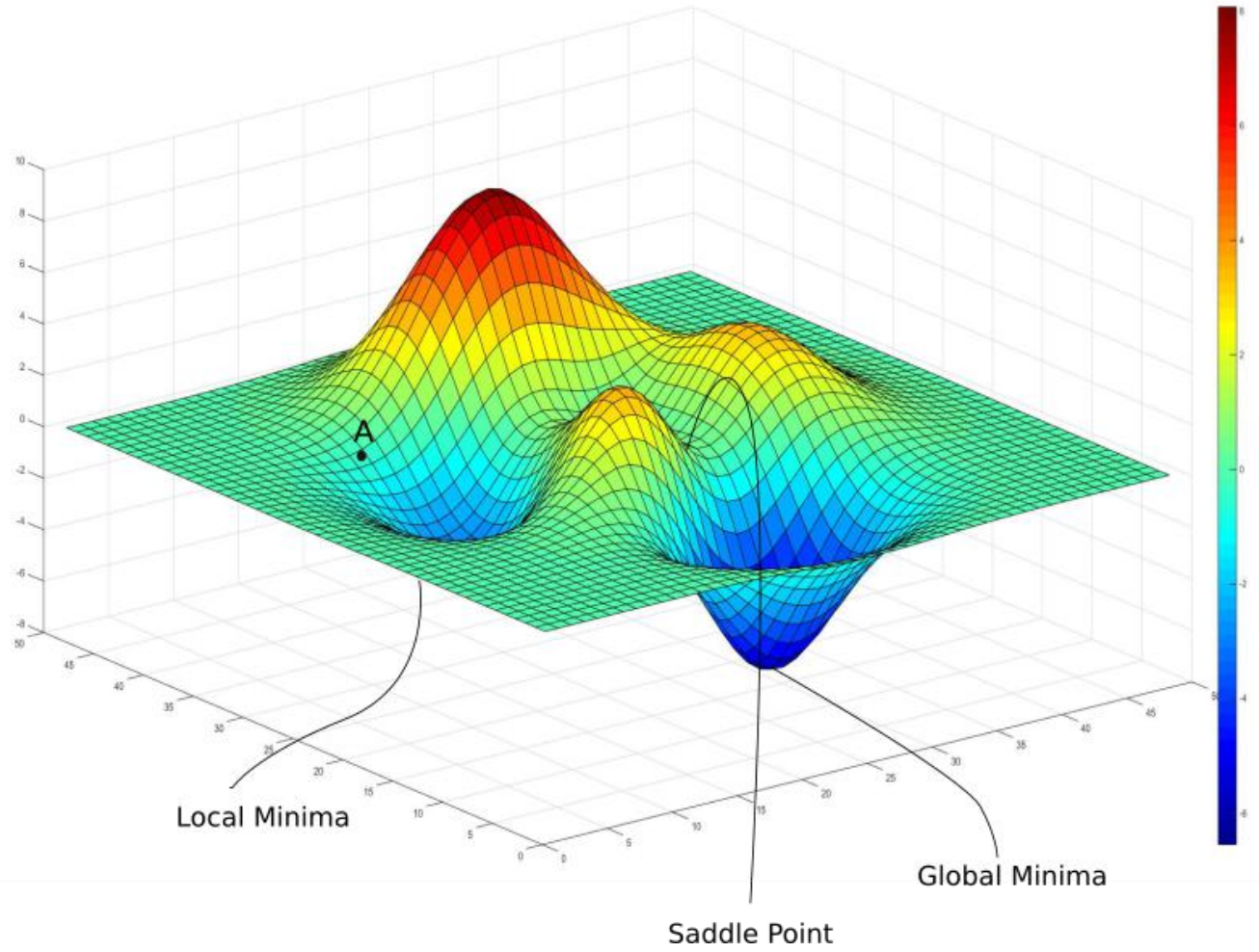


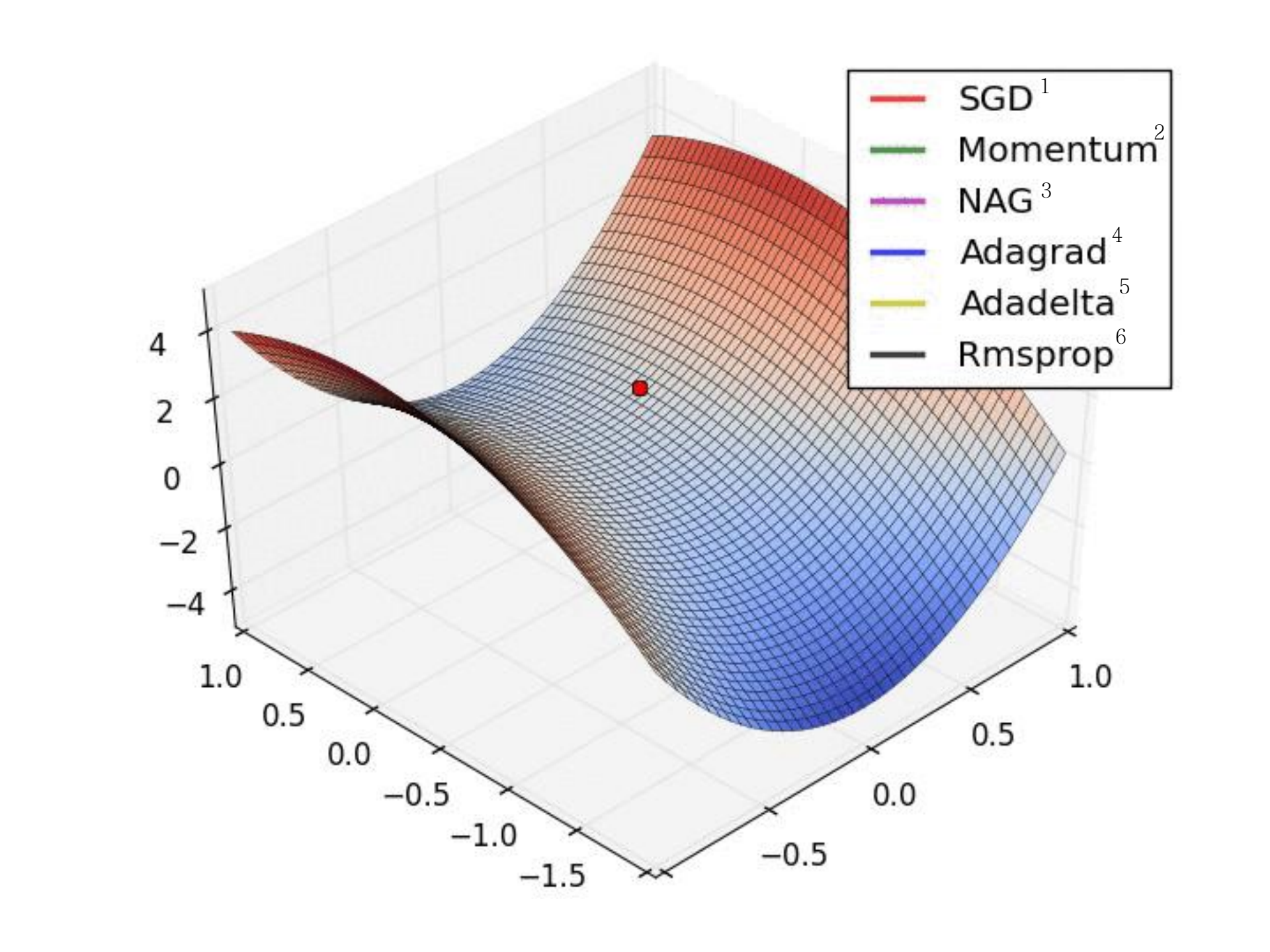

CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION


0 0 75 80 80
0 75 80 80 80
0 75 80 80 80
0 75 80 80 80
0 70 75 80 80
0 0 0 0 0
0 0 0 0 0
-1 -2 -1
-1 0 0
0 0 1
1 2 1
155

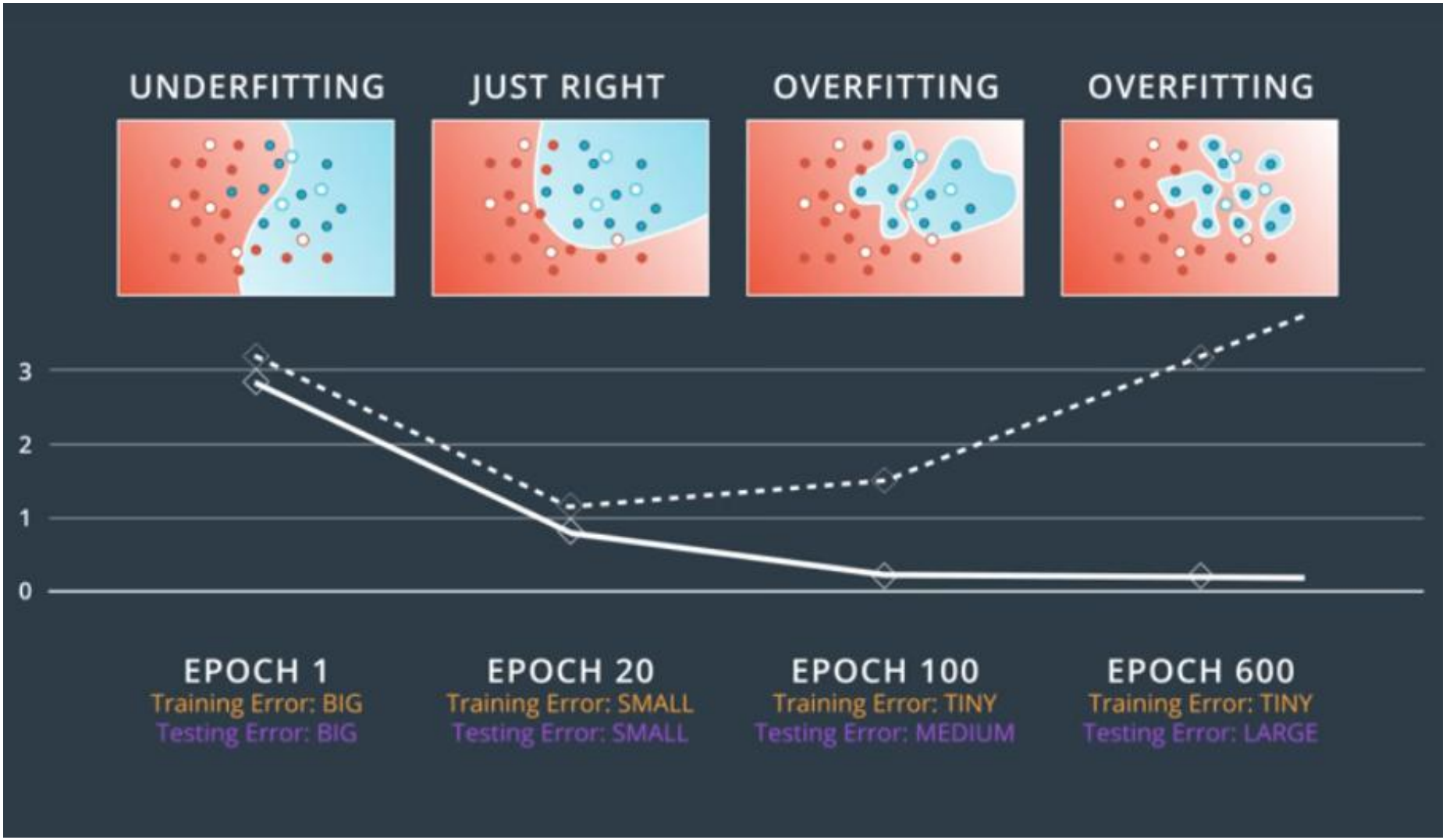

-4 5 3
5 -4 6
0 9 -4
2 0 -3
-2 4 1
3 0 2
MAX
9
-1 -3 0 7 -9 5
2 0 1 -2 5 3
3 -3 -9 3 2 1


p=0.5
hidden fc layer
dropout layer
input layer
output layer
Training time


A
Local Minima
Saddle Point
Global Minima


SGD [1]
Momentum [2]
NAG [3]
Adagrad [4]
Adadelta [5]
Rmsprop [6]
4
2
0
−2
−4
1.0
0.5
0.0
−0.5
−1.0
−1.5
1.0
0.5
0.0
−0.5

1	Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010 (pp. 177-186). Physica-Verlag HD.
2	Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
3	Su, W., Boyd, S. and Candes, E., 2014. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. In Advances in Neural Information Processing Systems (pp. 2510-2518).
4	Ogren, P., Fiorelli, E. and Leonard, N.E., 2004. Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment. IEEE Transactions on Automatic control, 49(8), pp.1292-1302.
5	Zeiler, M.D., 2012. ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701.
6	Hinton, G., Srivastava, N. and Swersky, K., 2012. Rmsprop: Divide the gradient by a running average of its recent magnitude. Neural networks for machine learning, Coursera lecture 6e.


UNDERFITTING
JUST RIGHT
OVERFITTING
OVERFITTING
3
2
1
0
EPOCH 1
Training Error: BIG
Testing Error: BIG
EPOCH 20
Training Error: SMALL
Testing Error: SMALL
EPOCH 100
Training Error: TINY
Testing Error: MEDIUM
EPOCH 600
Training Error: TINY
Testing Error: LARGE