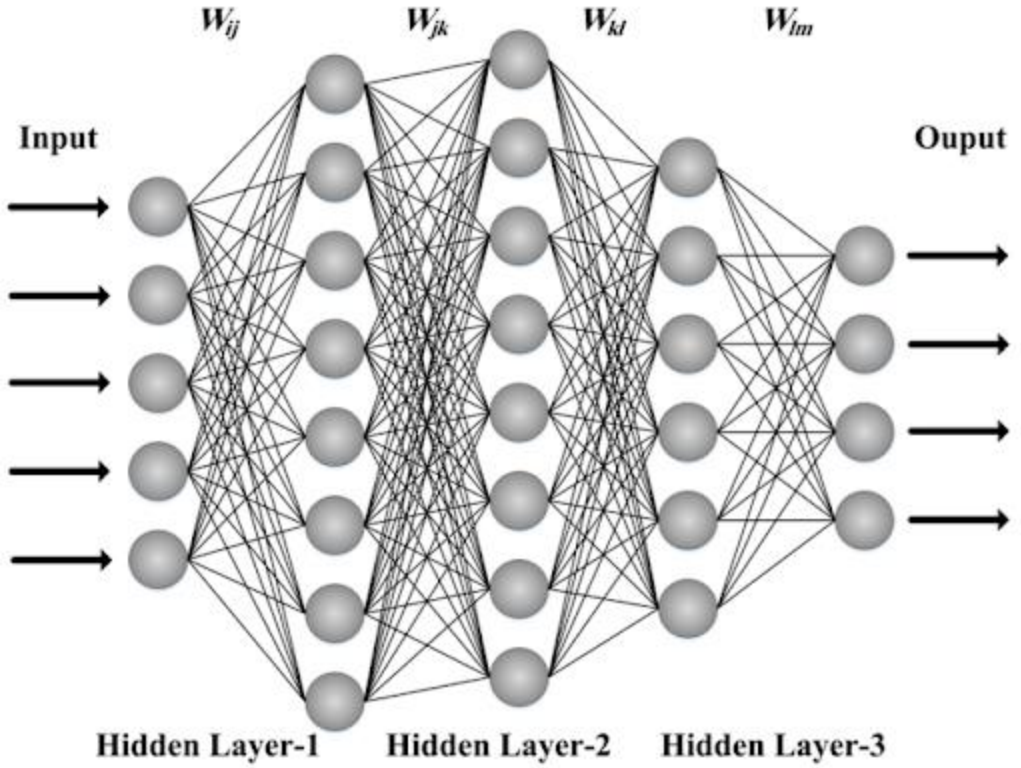
Weight Initialization

Weight Initialization

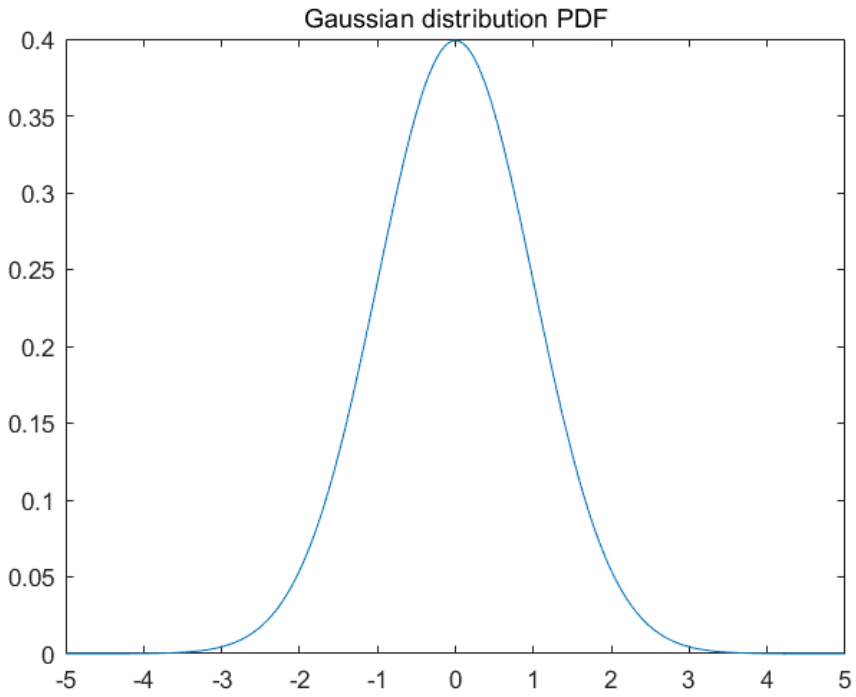
Weight Initialization

WI


W_{ij}
W_{jk}
W_{kl}
W_{lm}
Input
Ouput
Hidden Layer-1
Hidden Layer-2
Hidden Layer-3

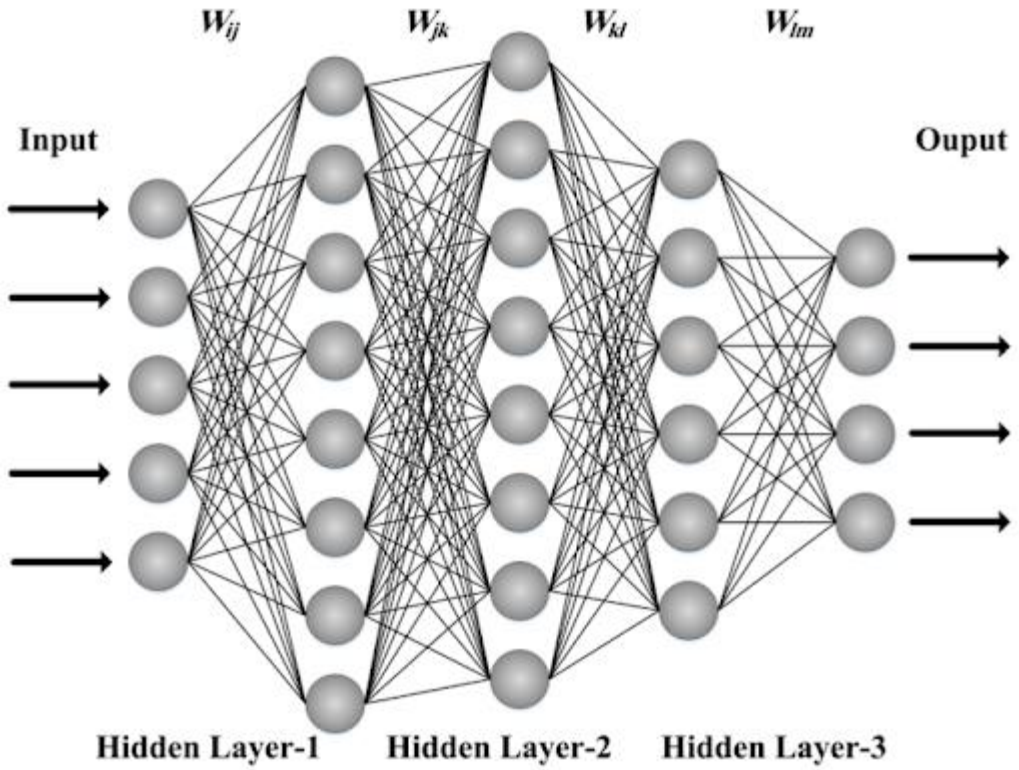
WI

Random function =>
Gaussian Distribution



The initial Weight
Gaussian Distribution

WI



The initial Sample
Gaussian Distribution

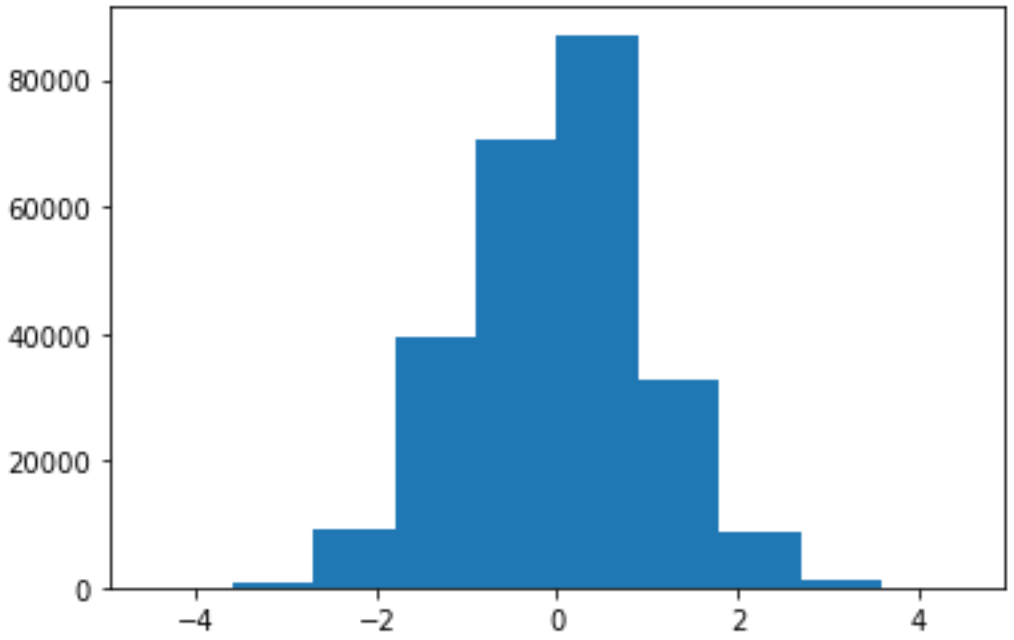
WI

Ex) 1-layer num of node is 500
Num of Sample is 500

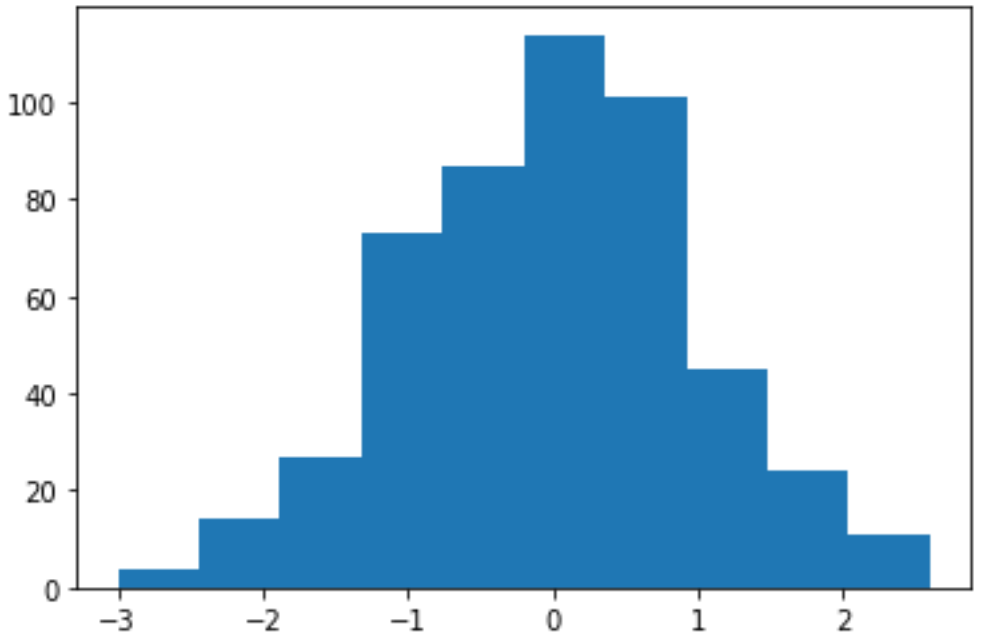
Matrix Calculation

500X1 Matrix
500X500 Matrix
Calculation

WI

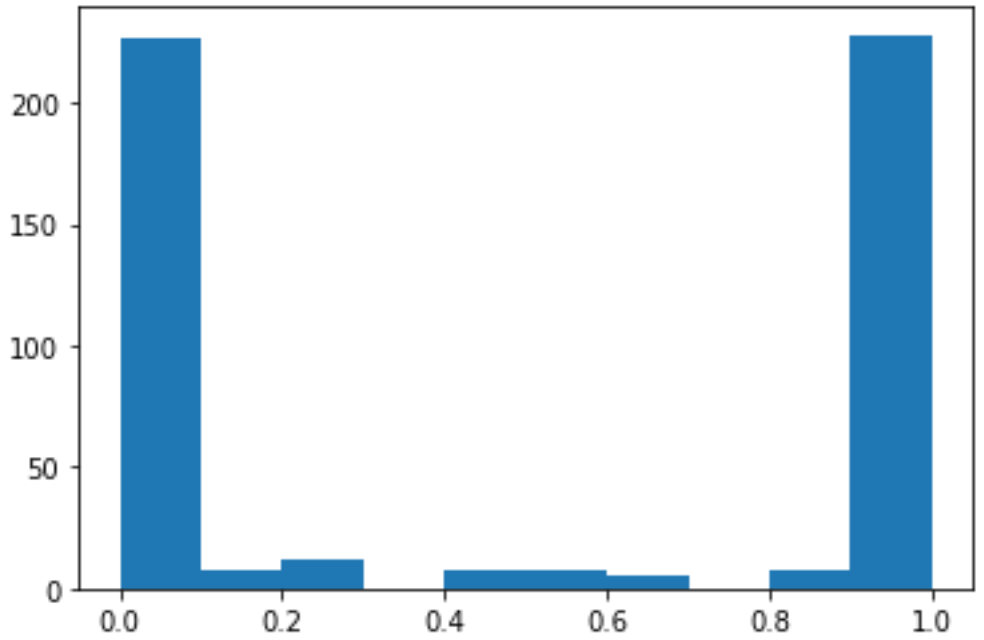


weight



Sample

WI

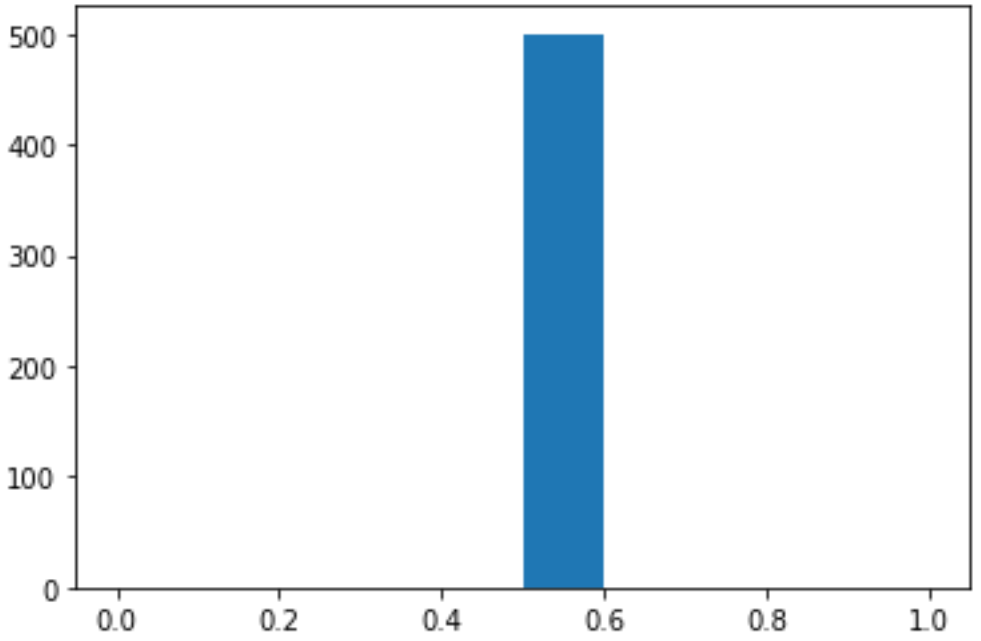


Sigmoid function

WI

LeCun
method 1

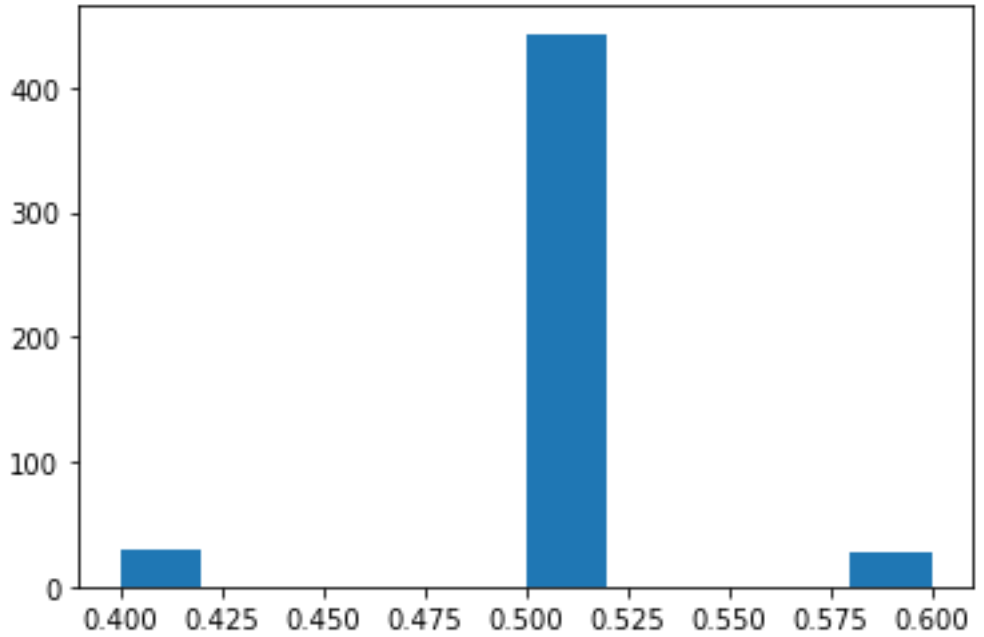
$$Var(W) = \sqrt{\frac{1}{n_{in}}}$$



WI

Xavier method 2

$$Var(W) = \sqrt{\frac{2}{n_{in} + n_{out}}}$$



WI

Paper

Understainding the difficulty of training deep feedforward neural networks

WI

ReLU Weigt initialization

He Initialization

$$Var(W) = \sqrt{\frac{2}{n_{in}}}$$

Bias Initialization

Bias => 0

ReLU => Bias 0.01

감사합니다

THANK YOU