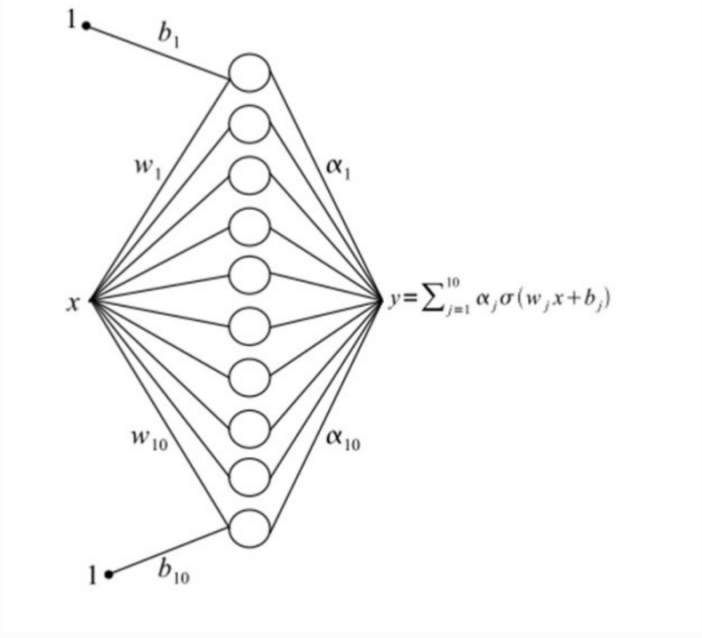
HW. (모든 node 연결, 각 neuron에 bias 존재 가정하에)

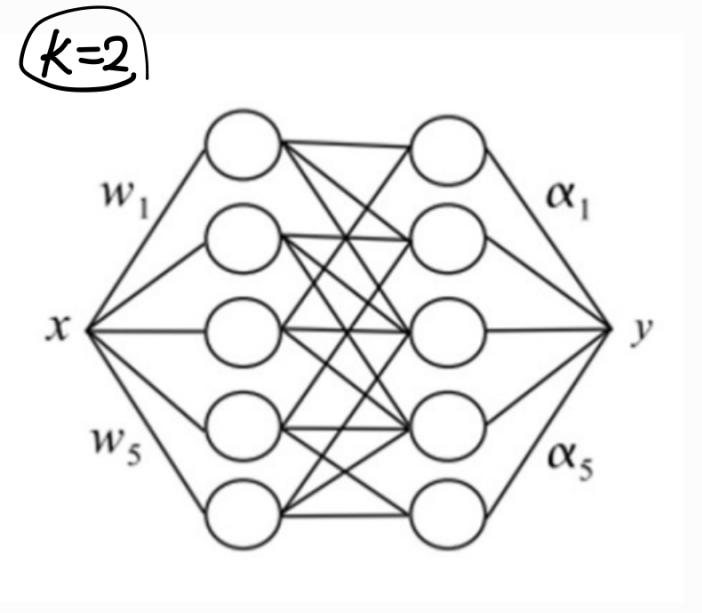
k=1



$10 + 10 + 10$

$= 30$

k=2

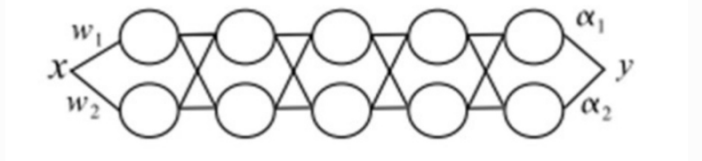


$5 + 5\times5 + 5 + 10$

$= 45$

$\rightarrow$ max

k=3



$2 + 2\times2 + 2\times2 + 2\times2$

$+ 2\times2 + 2 + 10$

$= 30$

Exercise 13.8.1 A feedforward neural network of type 784-200-100-50-10 is used to classify the MNIST data. Find the dimension of the associated output manifold. (784 is the input size and 10 represents the number of digit classes).

↳ parameter 수

$784 \times 200 + 200 \times 100 + 100 \times 50 + 50 \times 10$

$+ (200+100+50)$

bias

$= 182,650$

Let MNIST training data: 55000개
Let overfit: total parameter수>training data 수

Exercise 13.8.3 A one-hidden layer feedforward neural net, 784-N-10, is used to classify the MNIST data. Find the range of the number of hidden neurons, N, for which the network overfits the training data.

$784 \times N + N \times 10 + N > 55000$

$795N > 55000$

$N > 69.18$

If $N \geq 70$ then overfit

Exercise 13.8.4 A two-hidden layer feedforward neural net, 784-h-h-10, is used to classify the MNIST data. Find the range of the number h, for which the network overfits the training data.

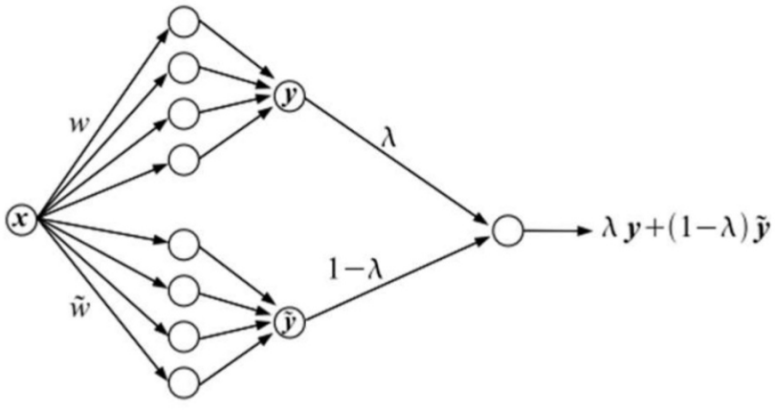
$784 \times h + h \times h + h \times 10 + 2h$

$= h^2 + 796h > 55000$

$h < -859.9567$ or $h > 63.9567$

If $h \geq 64$ then overfit

Exercise 13.8.10 Consider the model combination of two sigmoid neurons. Write the output of the combination and specify the dimension of the associated output manifold.



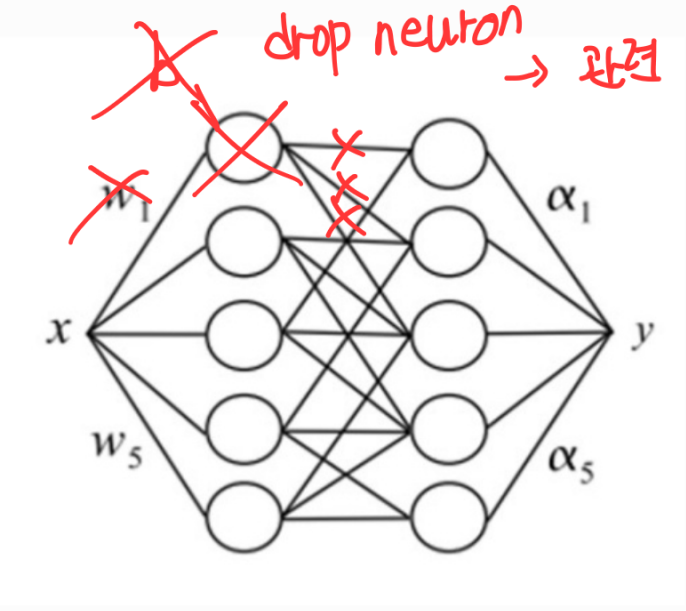
$y = \sigma(w_1 x + b_1)$

$\tilde{y} = \sigma(w_2 x + b_2)$

$\therefore$ output $= \lambda \sigma(w_1 x + b_1) + (1-\lambda)\sigma(w_2 x + b_2)$

$(\lambda, w_1, b_1, w_2, b_2) \rightarrow$ 5 dimension

Exercise 13.8.11 List a few effects of dropping neurons from a network on the associated output manifold.



drop neuron(s)

→ the number of parameters decreased

→ " means dimension of output manifold

→ we cosider (overfitted) when

dimension of output manifold > train data

(Or simply, we know that as the number of parameters increases, the risk of overfitting increases too)

→ dim of output manifold decreased,
then reduce overfitting.