DS 598 DEEP LEARNING

Sungjoon Park (BUID: U38522578)

Feb. 6th, 2024

Problem Set 4

1. (Problem 4.4)

$$\boldsymbol{h}_1 = a[\boldsymbol{\beta}_0 + \boldsymbol{\Omega}_0\boldsymbol{x}]$$
$$\boldsymbol{h}_2 = a[\boldsymbol{\beta}_1 + \boldsymbol{\Omega}_1\boldsymbol{h}_1]$$
$$\boldsymbol{h}_3 = a[\boldsymbol{\beta}_2 + \boldsymbol{\Omega}_2\boldsymbol{h}_2]$$
$$\boldsymbol{y} = \boldsymbol{\beta}_3 + \boldsymbol{\Omega}_3\boldsymbol{h}_3$$

$<$ From the input layer to the first hidden layer $>$
Input vector $\boldsymbol{x}$: 5×1
Weight matrix: $\boldsymbol{\Omega}_0$: 20×5
Bias vector: $\boldsymbol{\beta}_0$: 20×1
$\boldsymbol{h}_1$: 20×1

$<$ From the first hidden layer to the second hidden layer $>$
Weight matrix: $\boldsymbol{\Omega}_1$: 10×20
Bias vector: $\boldsymbol{\beta}_1$: 10×1
$\boldsymbol{h}_2$: 10×1

$<$ From the second hidden layer to the third hidden layer $>$
Weight matrix: $\boldsymbol{\Omega}_2$: 7×10
Bias vector: $\boldsymbol{\beta}_2$: 7×1
$\boldsymbol{h}_3$: 7×1

$<$ From the third hidden layer to the output layer $>$
Weight matrix: $\boldsymbol{\Omega}_3$: 4×7
Bias vector: $\boldsymbol{\beta}_3$: 4×1
$\boldsymbol{y}$: 4×1

If these functions are incorporated into one,

$$\boldsymbol{y} = \boldsymbol{\beta}_3 + \boldsymbol{\Omega}_3 a[\boldsymbol{\beta}_2 + \boldsymbol{\Omega}_2 a[\boldsymbol{\beta}_1 + \boldsymbol{\Omega}_1 a[\boldsymbol{\beta}_0 + \boldsymbol{\Omega}_0 \boldsymbol{x}]]]$$

2. (Problem 4.5)
   Since the depth of neural network is the number of its hidden layers and the width of it is the number of hidden units in each layer,
   Depth: 20
   Width: 30

3. (Problem 4.6)
   $<$ When the depth is increased by one $>$
   The number of weights between the input and the first hidden layers

$$1 \times 10 = 10$$

The number of weights between hidden layers

$$10 \times 10 \times (10 + 1) = 1100$$

The number of weights between the last hidden and output layers

$$10 \times 1 = 10$$

Total number of weights

$$10 + 1100 + 10 = 1120$$

$<$ When the width is increased by one $>$
The number of weights between the input and the first hidden layers

$$1 \times (10 + 1) = 11$$

The number of weights between hidden layers

$$(10 + 1) \times (10 + 1) \times 10 = 1210$$

The number of weights between the last hidden and output layers

$$(10 + 1) \times 1 = 11$$

Total number of weights

$$11 + 1210 + 11 = 1231$$

$\therefore$ Increasing the width by one generates more weights than increasing depth by the same number.

4. (Problem 4.10)

$<$ # of parameters between input and the first hidden layers $>$
- weights: D
- biases: $D+1$

$<$ # of parameters between hidden layers $>$
- weights: $D \times D \times (K-1)$
- biases: $D \times (K-1)$

$<$ # of parameters between the last hidden and output layers $>$
- weights: D
- biases: 1

$<$ Total number of parameters $>$

$$D+(D+1)+D\times D\times(K-1)+D\times(K-1)+D+1 = 3D+1+(K-1)D(D+1)$$