T1

Forward path

















end forward path

backward path

define 





















define 















T2

Size of A is 98\*98\*5

Size of B is 96\*96\*10

Size of C is 46\*46\*10

Size of D is 96\*96\*10

Size of E is 96\*96\*5

OT1

number of multiplication of A is

given input size 100\*100\*3

for each convolution neuron:

filter size and dept= 3\*3\*3 per round

we have to do: 98\*98 rounds

number of neuron: 5

So, the total number of multiplication of A is 3\*3\*3\*98\*98\*5 = 1296540

output size: 98\*98\*5

number of multiplication of B is

given input size 98\*98\*5

for each convolution neuron:

filter size and dept= 3\*3\*5 per round

we have to do: 96\*96 rounds

number of neuron: 10

So, the total number of multiplication of B is 3\*3\*5\*96\*96\*10 = 4147200

Total number of multiplication of path a to b is: 1296540+4147200 = 5443740

number of multiplication of D is

given input size 100\*100\*3

for each convolution neuron:

filter size and dept= 5\*5\*3 per round

we have to do: 96\*96 rounds

number of neuron: 10

So, the total number of multiplication of B is 5\*5\*3\*96\*96\*10 = 6912000

The parameter of A can be computed by:

given 3\*3 filter size and 3 input dimentions

each convolution neuron will have 3\*3\*3 = 27 weights

We have 5 filters so the total learnable parameters is 27\*5 = 135 + 5 biases = 140 parameters.

The parameter of B can be computed by:

given 3\*3 filter size and 5 input dimentions

each convolution neuron will have 3\*3\*5 = 45 weights

We have 10 filters so the total learnable parameters is 45\*10 = 450 + 10 biases = 460 parameters.

Total learnable parameters of path a to b is 140+460 = 600 parameters.

The parameter of D can be computed by:

given 5\*5 filter size and 3 input dimentions

each convolution neuron will have 5\*5\*3 = 75 weights

We have 10 filters so the total learnable parameters is 75\*10 + 10 biases = 760 parameters.

Total learnable parameters of path d is 760 parameters.

T3

In case that 







and compute



Hence, , when 

In case that 







and compute



Hence, , when 

Therefore, P(y=i)-1$, when i is the correct class and P(y=i)$, when i is not the correct class.

In other word, $P(y=i)-y\_i$