

CSE 676 Assignment 2, Part V

Part V.1: CNN

CSE 676 - A2 Part V

1. Input = 32×28 ,
10 filters, 5×5 , Stride = 1
 $(N - F) / \text{Stride} + 1$
 $= (32 - 5) / 1 + 1$
 $= 27 + 1$
 $= 28$ for height

$$\therefore \text{Its } 28 \times 24 \times 10$$

$$(28 - 5) / 1 + 1$$
$$= 23 + 1$$
$$= 24 \text{ for width}$$

2. - Each filter has size 5×5

- Input is RGB So has 3

$$5 \times 5 \times 3 = 75$$

- 10 filters total

$$(75 + 1) \times 10 = 760 \text{ parameters}$$

3. Padding = 1, 10 filters

- Means we add 1 pixels worth to each side,
So height & width ± 2

$$(32 - 5 + 2) / 1 + 1$$
$$= 29 + 1$$
$$= 30 \text{ height}$$

$$\therefore 30 \times 26 \times 10$$

$$(28 - 5 + 2) / 1 + 1$$
$$= 25 + 1$$
$$= 26 \text{ width}$$

4. - Greyscale is 1 instead of RGB's 3

- Still have 10 5×5 filters

$$(5 \times 5 + 1) \times 10 = 260 \text{ parameters}$$

(including biases)

5. Given the task and the requirement of probabilistic outputs, which activation function is most suitable for the output layer? Explain your choice and why other common activation functions are less appropriate in this scenario.

- Most Suitable activation function: Softmax
 - Because it would make sure all values are between 0 and 1 and sum to 1, giving us the probability for each of the 5 classes
- Why the others are less appropriate:
 - If we chose Sigmoid, then while we would get probabilities between 0 and 1, they would not necessarily sum to 1 (so its better for binary classification than multiclass)
 - If we chose ReLu, while it would set negative values to be 0, positive values would remain exactly the same, not giving us a probability distribution
 - If we chose Tanh, it would include values between -1 and 1, and negative values can't be used for probabilities
 - If we chose Linear activation, then the values would just stay the same

6. Prove that the activation function you chose in Task 5 is invariant to constant shifts in the input values. Mathematically demonstrate that adding a constant value to all input values will not change the resulting output probabilities. Explain the significance of shift invariance in this context.

6. $f(x) = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}$

- x is the input, so we're adding a constant C

$$\begin{aligned}
 g(x) &= \frac{e^{x_i + C}}{\sum_{j=1}^J e^{x_j + C}} \\
 &= \frac{e^{x_i} \cdot e^C}{\sum_{j=1}^J (e^{x_j} \cdot e^C)} \\
 &= \frac{e^{x_i} \cdot e^C}{e^C \cdot \sum_{j=1}^J e^{x_j}} \\
 &= \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}} \cdot \frac{\cancel{e^C}}{\cancel{e^C}} \\
 &= \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}
 \end{aligned}$$

\therefore Invariant to Constant Shifts