CS280 Fall 2022 Assignment 2
Part A

Convolutional Neural Nets

November 6, 2022

Name:Dai ZiJia

Student ID:2022233158

1. Convolution Cost (10 points)

Assume an input of shape $c_i \times h \times w$ and a convolution kernel of shape $c_o \times c_i \times k_h \times k_w$, padding of (p_h, p_w) , and stride of (s_h, s_w) .

- What is the computational cost (multiplications and additions) for the forward propagation?
- What is the memory footprint?

① Computation Cost is equal to output size multiply element-wise computation cost.

Output size:

$$size_{out} = C_o * \left(\frac{h-k_h+2P_h}{S_h}+1\right) * \left(\frac{w-k_w+2P_w}{S_w}+1\right)$$

Element-wise cost:

$$Cost_e = k_h * k_w + k_h * k * w = 2k_h * k_w$$

$$\text{Total cost} = size_o * cost_e$$

② Memory footprint include input m_i, output m_o kernel weight m_k and convolutional operation m_c

$$m_i = c_i * h * w$$
$$m_o = size_o$$
$$m_k = (C_i * k_h * k_w S + 1) * C_o$$
$$m_c = k_h * k_w + 1$$

$$\text{Total memory footprint} = m_i + m_o + m_k + m_c$$

2. Residual and Inception blocks (5 points)

What are the major differences between the Inception block and the residual block? After removing some paths in the Inception block, how are they related to each other?

Ans.

- Inception block uses multiple paths while residual block uses one single path with X.
- Thay can relate to others by adding a path connected each other like a residual block, may be there will need a 1x1 Conv.

3. Optimization (5 points)

Consider a simple multilayer perceptron with a single hidden layer of, say, d dimensions in the hidden layer and a single output. Show that for any local minimum there are at least $d!$ equivalent solutions that behave identically.

Ans. If one permutes the connections of the hidden layer ($d!$ ways to do that), and move and rename connections appropriately, then one effectively has the same MLP with the exact same minima, yet the configuration has changed (in a trivial sense). Thus there are at least $d!$ configurations only trivialy different with the exact same minima.

As the network is a MLP the equation would be

$$O_{hidden} = \sum_{i=1}^{d} w_{\pi_i} \cdot x + b_{\pi_i}$$

π_i is some order of the connections. For example $\pi_i = i$. And for d items there are $d!$ permutations thus $d!$ order functions $\pi(i) = \pi_i$.