

CS280 Fall 2021 Assignment 2

Part A

Convolutional Neural Nets

November 13, 2021

Name: FanYuxin

Student ID: 2020233216

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?

$$\begin{aligned} \text{multiplications}(c_i \times h \times w, c_o \times c_i \times k_h \times k_w) &= c_o \times \text{multiplications}(c_i \times h \times w, c_i \times k_h \times k_w) \\ \text{multiplications}(c_i \times h \times w, c_i \times k_h \times k_w) &= c_i \times \text{multiplications}(h \times w, k_h \times k_w) \\ \text{multiplications}(h \times w, k_h \times k_w) &= k_h \times k_w \times \text{outsize} \\ \text{outsize} &= \frac{h-k_h+2 \times p_h+s_h}{s_h} \times \frac{w-k_w+2 \times p_w+s_w}{s_w} \\ \text{total multiplications} &= c_o \times c_i \times k_h \times k_w \times \frac{h-k_h+2 \times p_h+s_h}{s_h} \times \frac{w-k_w+2 \times p_w+s_w}{s_w} \end{aligned}$$

$$\begin{aligned} \text{additions}(c_i \times h \times w, c_o \times c_i \times k_h \times k_w) &= c_o \times \text{additions}(c_i \times h \times w, c_i \times k_h \times k_w) \\ \text{additions}(c_i \times h \times w, c_i \times k_h \times k_w) &= c_i \times \text{additions}(h \times w, k_h \times k_w) + \text{outsize} \times (c_i - 1) \\ \text{additions}(h \times w, k_h \times k_w) &= (k_h \times k_w - 1) \times \text{outsize} \\ \text{total additions} &= c_o \times (c_i \times (k_h \times k_w - 1) \times \frac{h-k_h+2 \times p_h+s_h}{s_h} \times \frac{w-k_w+2 \times p_w+s_w}{s_w} + (c_i - 1) \times \frac{h-k_h+2 \times p_h+s_h}{s_h} \times \frac{w-k_w+2 \times p_w+s_w}{s_w}) \end{aligned}$$

Input: $c_i \times h \times w$

conv: $c_o \times c_i \times k_h \times k_w$

output: $\frac{h-k_h+2 \times p_h+s_h}{s_h} \times \frac{w-k_w+2 \times p_w+s_w}{s_w}$

memory footprint = $c_i \times h \times w + c_o \times c_i \times k_h \times k_w + \frac{h-k_h+2 \times p_h+s_h}{s_h} \times \frac{w-k_w+2 \times p_w+s_w}{s_w}$

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?

Inception uses multiple paths while residual block uses one single path with x . Remove two paths of inception block, we can get a block very similar to residual block. They both has two path, but residual block do nothing in a path while inception block do conv in both path. and residual add two path while inception concat paths.

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.

For a local minimum, to find equivalent local minimum, we can swap the weight of the d dimension hidden layer and swap the weight of d dim to 1 dim affine according to the swap in hidden layer.

Consider we have d weights and d positions, assign each weight to a position: there are totally $d!$ different assignments, they all have the same behavior.