Solution of Quiz4

Problem 1： Regularization

Batch Normalization: Show that the BN operator is differentiable. You can use the conclusion that elementary functions are differentiable.

solution:

BN formula:

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$$
$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2$$
$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$$
$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i)$$

$$\frac{\partial y_i}{\partial x_i} = \gamma \cdot \frac{\partial \hat{x}_i}{\partial x_i} = \gamma \cdot \left(\frac{1}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \right)$$

According to the chain rule, if a sub function (elementary function) is differentiable and the outer function is differentiable, then its composite function is also differentiable.

1. List the BN formula (including the rescaling part) or clearly describe the process of BN (inlcuding the rescaling part). (4)
2. Give the derivative function step by step or give it in chain rule form. (4)
3. Clearly prove and say the a sub function (elementary function) is differentiable and the outer function is differentiable to get the final conclusion. (2)

Quiz 3
Week 4, Oct/08/2019
CS 280: Fall 2019
Instructor: Xuming He

Name:
On your left:
On your right:

Problem 2. ResNet (10 points)
Consider a ResNet block, where *F* is a concatenation of two 3x3 conv + ReLU layers:

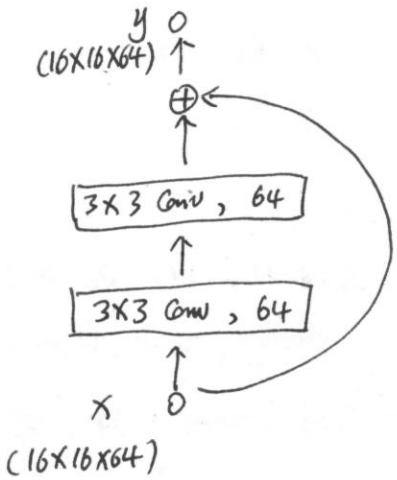
$$\mathbf{y} = F(\mathbf{x}) + \mathbf{x}$$

Assume x, intermediate layer and y are 16x16x64 feature maps. Modify *F* based on the bottleneck structure (Lecture 06, P46) such that the number of parameters of the block is reduced by 50-60%.

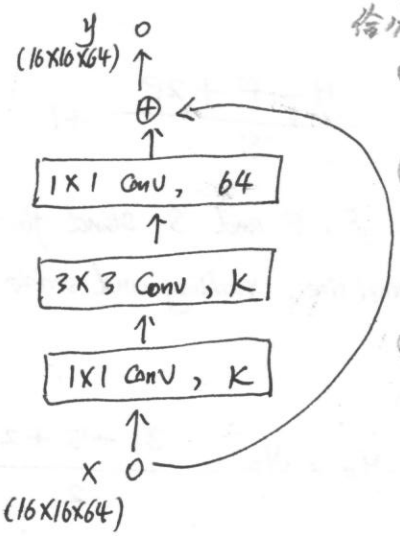
本题步骤：
画出B的网络结构图
写出A,B所用参数数量
求解二次不等式
给分：
① 画出或写出B的网络结构 2分
② 写对A,B所用参数数量 各2分
③ 写对不等式 2分
④ 求出结果且指定K范围为[51,57]的整数 2分

Solution:

A. 2 3X3 ConV + ReLU layers:



B. The bottleneck structure:



$A_{parameter\_number}$ =

$3 \times 3 \times 64 \times 64$
$+ 3 \times 3 \times 64 \times 64$

OR with bias

$= 36864 * 2 = 73728$

$(36864+64)*2 = 73856$

$B_{parameter\_number} = 1 \times 1 \times K \times 64$
$+ 3 \times 3 \times K \times K$
$+ 1 \times 1 \times K \times 64$
$= 9K^2 + 128K$

OR $9K^2 + 128K + 3*64$

To reduce the number of parameters by 50~60%,

$$40\% \leq \frac{9K^2+128K}{73728} \leq 50\%$$

OR

$$40\% \leq \frac{9K^2+128K+3*64}{73856} \leq 50\%$$

By solving above inequation,

$K \in [50.57, 57.28]$ (no bias)
OR $K \in [50.43, 57.17]$ (W. bias)
$\therefore \{51, 52, 53, 54, 55, 56, 57\}$