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)

Problem 2: ResNet

ResNet with bottleneck (L7 Page 42): Assume the input and output features are of size 10x10x18 and the conv kernel is 3x3. What is the maximal intermediate feature dimension, k, if you wish to reduce the number of multiplication operations at least by half? Provide the derivation of your solution.

Hint: first use 1x1 conv, k filters to generate 10x10xk feature maps, and then use 3x3 conv, k filters to generate 10x10xk feature maps and finally use 1x1 conv, 18 filters to generate output. Use padding if needed.

Solution: