P2. 1) $\dfrac{\partial Y_L}{\partial b_i} = \dfrac{\partial Y_L}{\partial Y_i} \dfrac{\partial Y_i}{\partial b_i} = \boxed{\dfrac{\partial Y_L}{\partial Y_{L-1}} \dfrac{\partial Y_{L-1}}{\partial Y_{L-2}} \cdots \dfrac{\partial Y_{i+1}}{\partial Y_i}} \cdot \dfrac{\partial Y_i}{\partial b_i}$

$$= \boxed{A_L \cdot \left(\text{diag}\left(\sigma'(\tilde{Y}_{L-1})\right) \cdot A_{L-1}\right) \cdot \cdots \cdot \left(\text{diag}(\sigma'(\tilde{Y}_{i+1}))A_{i+1}\right)} \text{diag}(\sigma'(\tilde{Y}_i))$$

$$\dfrac{\partial Y_L}{\partial A_i} = \text{diag}\left(\sigma'(\tilde{Y}_i)\right) \cdot \boxed{}^T Y_{i-1}^T$$

$\sigma'(x) = \sigma(x)(1-\sigma(x))$ 이고 $0 < \sigma(x) < 1$ 이고 $0 < \sigma'(x) \leq \frac{1}{4}$. 즉, $\text{diag}(\sigma'(\tilde{Y}_k))A_k$는
A_k의 not too large 혹은 small의 성질을 변화시키지 않는다.
$i = 1, \cdots, L$ 이므로 $L, L-1, \cdots, i+1$ 중 small인 matrix가 최소 1개 존재하게되고 $\boxed{}$ 값이 small이 되게되므로
$\dfrac{\partial Y_L}{\partial b_i}, \dfrac{\partial Y_L}{\partial A_i}$ 는 small

2) if $\tilde{Y}_j$ has large absolute value for some $j \in \{l+1, \cdots, L-1\}$

2)에서 $\text{diag}\left(\sigma'(\tilde{Y}_j)\right)$의 대각성분이 small 이므로

$\left(\because \sigma'(x) = \sigma(x)(1-\sigma(x)) \text{ 인데 } x \to \pm\infty \text{ 이면 } \sigma(x) \to 0 \right)$

1)과 똑같이 시고 small

P3. θ 1 는 계산.

form I

$$\theta' = \theta^0 - \alpha g^0 + \beta(\theta^0 - \theta^{-1}) = \theta^0 - \alpha g^0$$

form II

$$v' = g^0 + \beta v^0$$

$$\theta' = \theta^0 - \alpha v' = \theta^0 - \alpha g^0 - \alpha \beta v^0 = \theta^0 - \alpha g^0$$

} same

θ^{k-1}, θ^k 가 두 form 모두 동일로 가정하자.

form 1에서 $\theta^{k+1} = \theta^k - \alpha g^k + \beta(\theta^k - \theta^{k-1})$

form 2에서 $v^k = -\frac{1}{\alpha}(\theta^k - \theta^{k-1})$

$$v^{k+1} = g^k + \beta v^k = g^k - \frac{\beta}{\alpha}(\theta^k - \theta^{k-1})$$

$$\theta^{k+1} = \theta^k - \alpha v^{k+1} = \theta^k - \alpha \left(g^k - \frac{\beta}{\alpha}(\theta^k - \theta^{k-1})\right)$$

$$= \theta^k - \alpha g^k + \beta(\theta^k - \theta^{k-1})$$

두 form의 θ^{k+1}이 같다.

수학적귀납법에의해 모든 θ^k가 같다.

P4 One-based-index

・첫번째 Conv layer
  Conv2d(3,64,k=3,p=1)을 통과 했을때

   모든 k개 채널 X의 R,G,B 값 전부를 참조하며
  그 m(가로축)과 n(y좌표)는 $i-1 \leq m \leq i+1, j-1 \leq n \leq j+1$

・ReLu는 receptive field에 영향이 없다.

・두번째 Conv layer를 통과했을때 (Y_1) $Y_1[k,i,j]$는
   $i-2 \leq m \leq i+2, j-2 \leq n \leq j+2$, $1 \leq m \leq 224$, $1 \leq n \leq 224$, $m, n \in \mathbb{Z}$
  인 $X[[1,2,3],m,n]$ 의 영향을 받는다.

・y_2 (pool) 은 pooling이 되기전 y의 $(2i,2j),(2i-1,2j),(2i,2j-1),(2i-1,2j-1)$
  의 영향을 받으므로
   $2i-3 \leq m \leq 2i+2, 2j-3 \leq n \leq 2j+2$, $1 \leq m \leq 224$, $1 \leq n \leq 224$, $m, n \in \mathbb{Z}$ 인 $X[[1,2,3],m,n]$
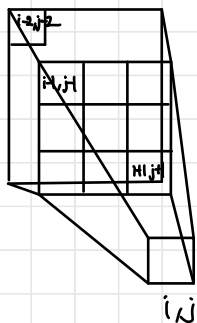 의 영향을 받는다.

・y_3이 되기전 Conv layer 2개에서는

   위의 1과정에 $i-2$를 넣은 것이 최소값, $i+2$를 넣은 것이 최댓값이 된다. j도 마찬가지.
   $2i-7 \leq m \leq 2i+6$, $2j-7 \leq n \leq 2j+6$, $1 \leq m \leq 224$, $1 \leq n \leq 224$, $m, n \in \mathbb{Z}$ 인 $X[[1,2,3],m,n]$
   pooling을 통과한 y_3에서는 y_2에서의 pooling 과 같은 논리로
   $4i-9 \leq m \leq 4i+6$, $4j-9 \leq n \leq 4j+6$, $1 \leq m \leq 224$, $1 \leq n \leq 224$, $m, n \in \mathbb{Z}$인 $X[[1,2,3],m,n]$
    가 된다.

P5. (i) 왼쪽: 1×1 conv: $(256 \times 1 \times 1) \times 128 + 128 = 32896$

3×3 conv: $(256 \times 3 \times 3) \times 192 + 192 = 442560$

5×5 conv: $(256 \times 5 \times 5) \times 96 + 96 = 614496$

total 1089952

오른쪽. 1×1 conv: $(256 \times 1 \times 1) \times 128 + 128 = 32896$

1×1 conv + 3×3 conv: $(256 \times 1 \times 1) \times 64 + 64 + (64 \times 3 \times 3) \times 192 + 192 = 127232$

1×1 conv + 3×3 conv: $(256 \times 1 \times 1) \times 64 + 64 + (64 \times 5 \times 5) \times 96 + 96 = 170144$

1×1 conv: $(256 \times 1 \times 1) \times 64 + 64 = 16448$

total: 346720

(ii) 왼쪽:

	곱셈	덧셈	activation function
1×1 conv:	$256 \times 32 \times 32 \times 128$ $= 33554432$	곱셈 값 + 32×32×128 (bias)	32×32×128
3×3 conv :	$256 \times 3 \times 3 \times 32 \times 32 \times 192$ $= 452984832$	곱셈값 + 32×32×192	32×32×192
5×5 conv :	$256 \times 5 \times 5 \times 32 \times 32 \times 96$ $= 129146600$	곱셈값 + 32×32×96	32×32×96
total	1115669864	1116110848	425984

오른쪽:

	곱셈	덧셈	activation function
1×1 conv	왼쪽과 동일	왼쪽과 동일	왼쪽과 동일
1×1 and 3×3	$256 \times 1 \times 1 \times 32 \times 32 \times 64$ $+ 64 \times 3 \times 3 \times 32 \times 32 \times 192$	곱셈값 + 32×32×64 $+ 32 \times 32 \times 192$	32×32×64 + 32×32×142
1×1 and 5×5	$256 \times 1 \times 1 \times 32 \times 32 \times 64$ $+ 64 \times 5 \times 5 \times 32 \times 32 \times 96$	곱셈값 + 32×32×64 $+ 32 \times 32 \times 96$	32×32×64 + 32×32×96
1×1 conv	$256 \times 1 \times 1 \times 32 \times 32 \times 64$	곱셈값 + 32×32×64	32×32×64
total	354418688	354910208	622592