Problem1.

(a) $V_\beta(x) = \frac{1}{\beta}\log\sum_{i=1}^{n} e^{\beta x_i}$

$x_1,\cdots,x_n$ 중 가장 큰 값을 x_k라하자

$$V_\beta(x) = \frac{1}{\beta}\log e^{\beta x_k}\left(\frac{e^{\beta x_1}+\cdots+e^{\beta x_n}}{e^{\beta x_k}}\right) = x_k + \log\left(e^{\beta(x_1-x_k)}+\cdots+e^{\beta(x_n-x_k)}\right)^{\frac{1}{\beta}}$$

$A_\beta(x) = \log\left(e^{\beta(x_1-x_k)}+\cdots+e^{\beta(x_n-x_k)}\right)^{\frac{1}{\beta}}$ 라 하면, $x_i \le x_k$, $i=1,\cdots n$ 이고 $\beta>0$이면

$$0 \le e^{\beta(x_i-x_k)} \le 1 \quad \text{, 특히 } i=k\text{면 } e^{\beta(x_k-x_k)}=1$$

즉, $1 \le \left(e^{\beta(x_1-x_k)}+\cdots+e^{\beta(x_n-x_k)}\right) \le n$

$$1 \le \left(e^{\beta(x_1-x_k)}+\cdots+e^{\beta(x_n-x_k)}\right)^{\frac{1}{\beta}} \le n^{\frac{1}{\beta}}$$

$\beta\to\infty$ 이면

$$\left(e^{\beta(x_1-x_k)}+\cdots+e^{\beta(x_n-x_k)}\right)^{\frac{1}{\beta}} \to 1 \text{ 이므로 } A_\beta(x)\to 0$$

$\therefore$ $\beta\to\infty$이면 $V_\beta(x)\to x_k = \max\{x_1,\cdots,x_n\}$

(b) $\nabla V_1 = \frac{1}{e^{x_1}+\cdots+e^{x_n}}\begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_n}\end{bmatrix}$ softmax 함수와 동일하다.

(c) $\nabla V_\beta(x) = \frac{1}{e^{\beta x_1}+\cdots+e^{\beta x_n}}\begin{bmatrix} e^{\beta x_1} \\ \vdots \\ e^{\beta x_n}\end{bmatrix}$

$i\ne i_{max}$인 index: $\frac{e^{\beta(x_i-x_i)}}{\sum_{l=1}^{n} e^{\beta(x_l-x_i)}} = \frac{1}{\sum_{l=1}^{n} e^{\beta(x_l-x_i)}}$

$x_i < x_{i_{max}}$ 이므로 $\beta\to\infty$이면 0으로 간다.

$i=i_{max}$인 index $\frac{e^{\beta(x_{i_{max}}-x_{i_{max}})}}{\sum_{l=1}^{n} e^{\beta(x_l-x_{i_{max}})}} = \frac{1}{e^{\beta(x_1-x_{i_{max}})}+\cdots+1+\cdots+e^{\beta(x_n-x_{i_{max}})}}$

$\beta\to\infty$ 이면 1로 간다.

$\therefore$ $\nabla V_\beta(x)\to e_{i_{max}}$ as $\beta\to\infty$

P2.

conv1 에서	$(64\times55^2)\times(3\times11^2)$	
conv2	$192\times27^2\times64\times5^2$	
conv3	$384\times13^2\times192\times3^2$	655566528
conv4	$256\times13^2\times384\times3^2$	
conv5	$256\times13^2\times256\times3^2$	
linear1	9216×4096	
linear2	4096×4096	58621952
linear3	4096×1000	

Problem 4.

HW4 에서와 똑같이

$$\frac{\partial y_L}{\partial y_{L-1}} = A_{w_L}, \quad \frac{\partial y_\ell}{\partial y_{\ell-1}} = \operatorname{diag}\left(\sigma'(A_{w_\ell} y_{\ell-1} + b_\ell 1_{n_\ell})\right) A_{w_\ell}$$

$$\frac{\partial y_\ell}{\partial b_\ell} = \operatorname{diag}\left(\sigma'(A_{w_\ell} y_{\ell-1} + b_\ell 1_{n_\ell})\right) 1_{n_\ell}$$

$$\frac{\partial y_L}{\partial b_\ell} = \frac{\partial y_L}{\partial y_\ell}\frac{\partial y_\ell}{\partial b_\ell} = \frac{\partial y_L}{\partial y_\ell} \operatorname{diag}\left(\sigma'(A_{w_\ell} y_{\ell-1} + b_\ell 1_{n_\ell})\right) 1_{n_\ell} = v_\ell 1_{n_\ell}$$

$$(y_\ell)_i = \sigma\left(\sum_{j=i}^{i+f_\ell-1} (w_\ell)_{j-i} (y_{\ell-1})_j + b_\ell\right)$$

$$\frac{\partial y_\ell}{\partial (w_\ell)_k} = \operatorname{diag}\left(\sigma'(A_{w_\ell} y_{\ell-1} + b_\ell)\right) \begin{bmatrix} (y_{\ell-1})_k \\ \vdots \\ (y_{\ell-1})_{n_\ell+k-1} \end{bmatrix}$$

$$\frac{\partial y_L}{\partial (w_\ell)_k} = v_\ell \begin{bmatrix} (y_{\ell-1})_k \\ \vdots \\ (y_{\ell-1})_{n_\ell+k-1} \end{bmatrix}$$

$$\frac{\partial y_L}{\partial (w_\ell)} = v_\ell \begin{bmatrix} (y_{\ell-1})_1 & (y_{\ell-1})_2 & \cdots & (y_{\ell-1})_{f_\ell} \\ \vdots & \vdots & \ddots & \vdots \\ (y_{\ell-1})_{n_\ell} & & & (y_{\ell-1})_{n_\ell-1+f_\ell} \end{bmatrix}$$

이 연산은 $y_{\ell-1}$과 v_ℓ^T 와의 Conv 연산을 Transpose 한 것과 동일하다.

$$\therefore \frac{\partial y_L}{\partial (w_\ell)} = \left(C_{v_\ell^T} y_{\ell-1}\right)^T$$