1. Softmax

1. $$\text{softmax}(x) = \text{softmax}(x+c)$$

(proof) $\iff$ $\text{softmax}(x)_i = \text{softmax}(x_i + c)$ for all $\hat{n}$

(좌변) $= \dfrac{e^{x_i}}{\sum_j e^{x_j}}$

(우변) $= \dfrac{e^{(x_i + c)}}{\sum_j e^{(x_j + c)}} = \dfrac{e^{x_i} \cdot e^c}{\sum_j (e^{x_j} \cdot e^c)} = \dfrac{e^c \cdot e^{x_i}}{e^c \cdot \sum_j e^{x_j}} = \dfrac{e^{x_i}}{\sum_j e^{x_j}}$

$\therefore$ (좌변) $=$ (우변)

2. 만약 x vector의 element 값이 매우 큰 경우,
   $\exp(x)$ 는 최댓값 처리에 overflow 발생 가능.
   x 값이 매우 작은 경우도 $\exp(x) \to 0$ 분모 0으로 underflow 가능
   따라서 $\text{softmax}(x)$ 대신 $\text{softmax}(x - \max_i x_i)$ 사용.
   (Details : DL 78 page 참고)