TAREA APR

Sea $c = y_n$; podemos escribir:

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial(-\log p(y_n|\mu_n))}{\partial W^t} = \frac{1}{N}\sum_{n=1}^{N}\frac{\partial\left(-\log\left(\frac{e^{a_{nc}}}{\sum_{c'=1}^{C} e^{a_{nc'}}}\right)\right)}{\partial W^t}.$$

Ahora bien, esto es para n muestras.

Supongamos $n=1$:

$$= \boxed{\frac{\delta\left(-\log\left(\frac{e^{a_c}}{\sum_{c'=1}^{C} e^{a_{c'}}}\right)\right)}{\delta W^t}};$$

A partir de esto, podemos aplicar propiedades de log y obtener una expresión simplificada.

$$= \frac{\delta\left(-\log\left(\frac{\sum_{c'=1}^{C} e^{a_{c'}}}{e^{a_c}}\right)\right)}{\delta W^t} = \frac{\delta\left(\log\left(\sum_{c'=1}^{C} e^{a_{c'}}\right) - e^{a_c}\right)}{\delta W^t}$$

Ahora calculamos las derivadas respecto de las C filas de W^t.

Llamaremos W_j^t al vector de la fila j de W^t.

$$\frac{\delta\left(\log\left(\sum_{c'=1}^{C} e^{a_{c'}}\right) - e^{a_c}\right)}{dW_j^t} = \boxed{\underset{①}{\frac{\delta\left(\log\left(\sum_{c'=1}^{C} e^{a_{c'}}\right)\right)}{dW_j^t}}} - \boxed{\underset{②}{\frac{\delta(a_c)}{dW_j^t}}}$$

Calculamos las derivadas:

$$① = \frac{\delta\left(\log\left(\sum_{c'=1}^{C} e^{a_{c'}}\right)\right)}{dW^t} = \frac{e^{a_j}\cdot\vec{X}}{\sum_{c'=1}^{C} e^{a_{c'}}} = \boxed{\mu_j \cdot X}$$

$$② = \frac{\delta(a_c)}{dW_j^t} = \frac{\delta(W_c^t \cdot \vec{x})}{W_j^t} = \mathbb{I}(c=j)\cdot\vec{x}$$

$\mathbb{I}(x) \rightarrow$ devuelve 1 si x = True y 0 en otro caso. Es decir $\mathbb{I}(c=j) = 1$ si $c=j$.

Esto es porque W_j solo se tiene en cuenta para a_c si $c=j$, para que en ese caso la derivada sea $\vec{x}$ o 0, en otro caso.

Sustituimos:

$$\mu_j \cdot \vec{x} - \mathbb{I}(c=j)\cdot\vec{x} = (\mu_j - \mathbb{I}(c=j))\cdot\vec{x}$$

La matriz resultante será:

$$\begin{pmatrix} (\mu_i - \mathbb{I}(c=1))\cdot\vec{x}^t \\ (\mu_i - \mathbb{I}(c=2))\cdot\vec{x}^t \\ \vdots \\ (\mu_c - \mathbb{I}(c=C))\cdot\vec{x}^t \end{pmatrix}.$$

Considerando $\vec{y}$ codificado en one-hot:

$$(\mu - \vec{y})\cdot\vec{x}^t.$$

Ahora con N muestras:

$$\boxed{\frac{1}{N}\sum_{n=1}^{N}(\mu_n - \vec{y}_n)\cdot\vec{x}_n^t} \quad C.Q.D$$

$$\Downarrow$$

$$\boxed{\frac{1}{N}\sum_{n=1}^{N}\vec{x}_n(\mu_n - \vec{y}_n)} \quad (\text{Traspuesta} \rightarrow D\times C \rightarrow C\times D)$$