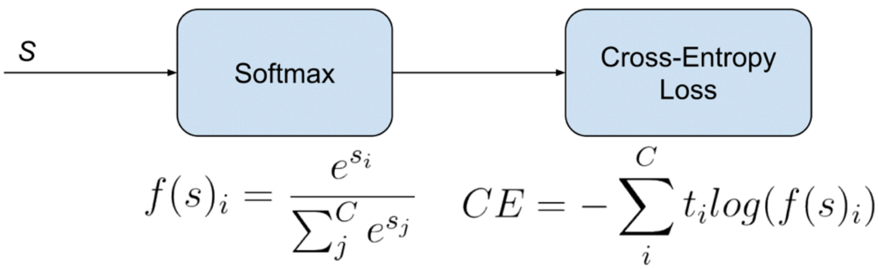
```
**************************************
**TALLER # 2 Categorical Cross Entropy**
**************************************
```

Elaborado por:

- Juanita Piraban Barbosa - 201216313
- Lorena Morales Rodriguez - 202027957
- Alexander Zapata Galindo - 201425426
- Jaime Humberto Trujillo Perea - 201920366
- Alejandro Barinas Guio - 201628859

De acuerdo a lo visto en clase y con una búsqueda en la literatura relevante, establecer una expresión para el gradiente de la función de costo, Co, considerando como función de activación en la salida, la función softmax y la entropía cruzada categórica como función de pérdida.



Tomado de: https://gombru.github.io/2018/05/23/cross_entropy_loss/ (https://gombru.github.io/2018/05/23/cross_entropy_loss/)

Reescribiendo la anterior expresión en términos de ‘Y’ (reemplanzando ‘s’ y ‘t’), se tiene:

$$f(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{c} e^{y_j}}, \text{ función de activación Softmax}$$

$$CE \ = \ -\sum_{i=1}^{c} y_i \ log \ \hat{y}_i, \text{ función de pérdida categorical crossentropy}$$

Solución:

Sean:

$Y = [\ 0\ 0\ 0\ 1\ ...\ 0\ ]$: vector de entrada

$\hat{Y} = [\ \hat{y}_1\ \hat{y}_2\ \hat{y}_3\ \hat{y}_4\ \ldots\ \ \hat{y}_c\ ]$: vector de salida donde c es el # de clases

$C_0 = -\sum_{i=1}^{c} y_i \ log\ \hat{y}_i$,donde $y_i \in Y,\ \hat{y}_i \in \hat{Y}$: función categórica crossentropy de costo

Como $y_p = 1$, es cierto para la clase p y $y_i = 0$ para i diferente de p, entonces:

$$C_0 = -[\ y_1\ log\ \hat{y}_1 + y_2\ log\ \hat{y}_2 +\ \ldots\ + y_p\ log\ \hat{y}_p +\ \ldots\ + y_c\ log\ \hat{y}_c]$$
$$= -[0\ +\ 0\ +\ \ldots\ +\ log\ \hat{y}_p\ +\ \ldots\ +\ 0]$$
$$= -\ log\ \hat{y}_p\ (1)$$

Es decir, solo la clase **p** contribuye a la función de pérdidas.

Ahora, como $\sigma(y_i) = e^{y_i} / \sum_{j=1}^{c} e^{y_j}$ (2) es la funcion de activación softmax, entonces reemplazando (2) en (1) :

$$C_0 = -\ log(e^{\hat{y}_p} / \sum_{j=1}^{c} e^{\hat{y}_j})$$

Derivando C_0 con respecto a $\hat{y}_p$ clase positiva:

$$\frac{\partial C_0}{\partial \hat{y}_p} = \frac{-1}{e^{\hat{y}_p} / \sum_{j=1}^{c} e^{\hat{y}_j}} * [\frac{\sum_{j=1}^{c} e^{\hat{y}_j} e^{\hat{y}_p} - e^{\hat{y}_p} e^{\hat{y}_p}}{(\sum_{j=1}^{c} e^{\hat{y}_j})^2}]$$
$$= \frac{-\sum_{j=1}^{c} e^{\hat{y}_j}}{e^{\hat{y}_p}} * \frac{e^{\hat{y}_p}}{\sum_{j=1}^{c} e^{\hat{y}_j}} * \frac{\sum_{j=1}^{c} e^{\hat{y}_j} - e^{\hat{y}_p}}{\sum_{j=1}^{c} e^{\hat{y}_j}}$$
$$= -[\frac{\sum_{j=1}^{c} e^{\hat{y}_j} - e^{\hat{y}_p}}{\sum_{j=1}^{c} e^{\hat{y}_j}}]$$
$$= \frac{e^{\hat{y}_p}}{\sum_{j=1}^{c} e^{\hat{y}_j}} - 1\ \ (3)$$

Ahora, derivando C_0 con respecto a algún $\hat{y}_n$ (clase negativa):

$$\frac{\partial C_0}{\partial \hat{y}_n} = \frac{-1}{e^{\hat{y}_p} / \sum_{j=1}^{c} e^{\hat{y}_j}} * [\frac{\sum_{j=1}^{c} e^{\hat{y}_j}(0) - e^{\hat{y}_p} e^{\hat{y}_n}}{(\sum_{j=1}^{c} e^{\hat{y}_j})^2}]$$
$$= \frac{-\sum_{j=1}^{c} e^{\hat{y}_j}}{e^{\hat{y}_p}} * \frac{-e^{\hat{y}_p} e^{\hat{y}_n}}{(\sum_{j=1}^{c} e^{\hat{y}_j})^2}$$
$$= \frac{e^{\hat{y}_n}}{\sum_{j=1}^{c} e^{\hat{y}_j}}\ \ (4)$$

En conclusión, la expresiones (3) y (4) corresponden al gradiente de una función de costo categorical crossentropy, con una función de activación Softmax.