Regresión Softmax

La regresión Softmax es una generalización de la regresión logística para clasificación multiclase. También se conoce como Multinomial Logistic Regression.

La idea es tener un puntaje para cada clase $c \in C$, para después estimar la probabilidad de pertenencia a cada clase usando la función Softmax.

Puntaje de una observación en la clase c:

$$s_c(x) = \beta_c^T X$$

(Ojo: tenemos un vector β_c para cada clase)

Diremos que Θ es la matriz que tiene cada β_c como una fila. Ahora la probabilidad de que una observación x pertenezca a la clase c es:

$$\boxed{\hat{p}_c = \sigma(s(x))_c = \frac{e^{s_c(x)}}{\sum_{j \in C} e^{s_j(x)}}}$$

y luego nuestra predicción $\hat{y}$ es:

$$\boxed{\hat{y} = \underset{c}{\arg\max}\ \sigma(s(x))_c}$$

Ahora, ¿cómo entrenamos este modelo? como función objetivo vamos a usar la función Cross Entropy

$$J(\Theta) = -\frac{1}{n} \sum_{i=1}^{n} \left[\sum_{c=1}^{C} \left[y_i^c \log \left(\hat{p}_c^{\,i} \right) \right] \right]$$

Donde:

- Tenemos n observaciones.
- Son C clases posibles.
- y_i^c es la probabilidad que queremos que sea asignada a la clase c en la observación i.
- $\hat{p}_c^{\,i}$ es la probabilidad que se predice para la clase c en la observación i.

Ojo! y_i^c tiende a ser 1 o 0, ya que deseamos tener probabilidad 1 en la clase correcta y 0 en los demás.

Así, si el número de clases es 2, volvemos a la función objetivo de la regresión logística.

Esta función objetivo castiga cuando tenemos una probabilidad baja para la clase que correspondía a la respuesta correcta.

Finalmente, el gradiente para cada clase es:

$$\nabla_{\beta^c} J(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{p}_c^{\,i} - y_i^c \right) x_i$$

Así encontramos el óptimo para cada β^c y así construir Θ.