Ciencia de Datos y Lenguaje Natural

Regresión Logística

Grupo PLN - INCO

Universidad de la República

Agenda

- Regresión logística, definición
- Análisis de sentimientos
- Atributos o variables de entrada
- ➔ Función de clasificación
- Método de entrenamiento
- Métodos de evaluación

Función de clasificación

Calculando atributos a partir del texto de entrenamiento, nos queda la ecuación de una recta en un espacio n-dimensional.

- $z = w_1x_1 + w_2x_2 + ... + w_nx_n + b$
- z $=$ $\mathbf{w.x}+$ b , notación vectorial con producto interno
- Los parámetros son los w_i y b.
- Deseamos tener 2 valores posibles de salida: positivo y negativo.
- Pero z varía en todo el dominio de los reales.

Modelo probabilístico

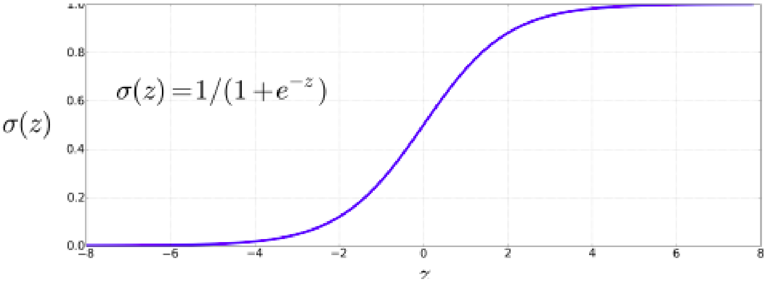
$$z = w_1x_1 + w_2x_2 + ... + w_nx_n + b$$

Mapeamos z en una variable que varíe en el intervalo [0,1]

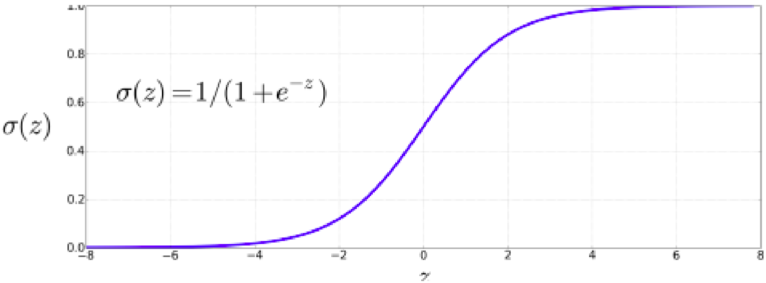
$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+\exp(-z)}$$

Función sigmoide

$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+\exp(-z)}$$



Propiedades de la sigmoide



1. $\sigma(z)$ varía en el intervalo [0,1], para z variando en $(-\infty, \infty)$
2. $1 - \sigma(z) = \sigma(-z)$
3. Es derivable, la derivada es muy fácil de calcular.
   $\sigma(z)' = \sigma(z).(1 - \sigma(z))$

Función de probabilidad para el clasificador

- $\sigma(z)$ es una función de probabilidad plausible para nuestro clasificador.
- $p(z = 1) = \sigma(z)$
- $p(z = 0) = 1 - \sigma(z)$

Transformando una probabilidad en un clasificador

$\hat{y}$ denota el valor que calcula el clasificador

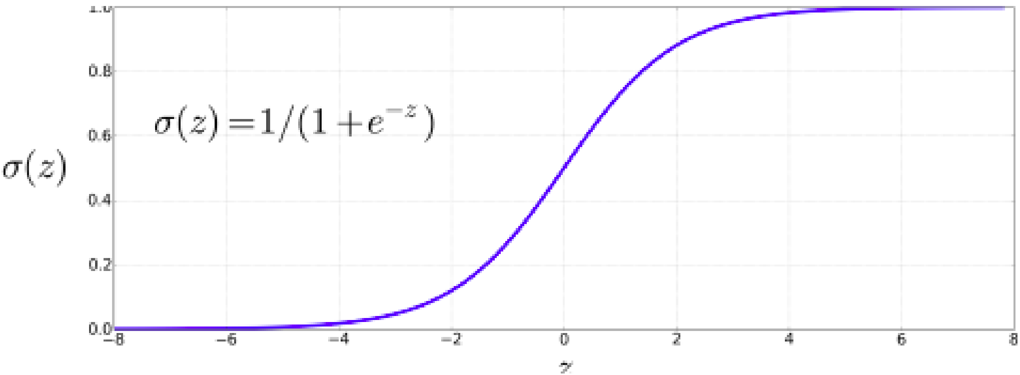
$$\hat{y} = \begin{cases} 1 \text{ si } p(y = 1|x) > 0{,}5 \\ 0 \text{ en caso contrario} \end{cases}$$

Transformando una probabilidad en un clasificador

ŷ denota el valor que calcula el clasificador

$$\hat{y} = \begin{cases} 1 \text{ si } p(y = 1|x) > 0{,}5 \text{ si } \mathbf{w.x} + \text{b} > 0 \\ 0 \text{ en caso contrario} \end{cases}$$

Transformando una probabilidad en un clasificador



$\sigma(0) = 0{,}5$

Agenda

- Regresión logística, definición
- Análisis de sentimientos
- Atributos o variables de entrada
- Función de clasificación
- ➔ Método de entrenamiento
- Métodos de evaluación

Función de pérdida

Conocemos el valor de la variable de salida y t.q. $y = \mathbf{w.x}+$ b en cada ejemplo de entrenamiento.

Deseamos que nuestro sistema calcule el estimado de y, $\hat{y}$ que mejor aproxime a y

O sea, minimizamos la distancia entre y y $\hat{y}$

1. Necesitamos una función de distancia, o función de pérdida (*loss function*)
2. El algoritmo que minimiza la distancia es descenso por gradiente

Estimador de máxima verosimilitud

Sea $L_{CE}(\hat{y}, y)$ la función de pérdida.

Los parámetros a determinar son los elementos del vector $\mathbf{w}$ y b

Los elegimos de modo que se maximice la probabilidad de los valores y verdaderos dados los valores de los atributos de entrada $\mathbf{w}$ - Estimador de máxima verosimilitud.

Entropía cruzada como función de pérdida

- Deseamos maximizar $p(y|x) = \hat{y}^y(1-\hat{y})^{1-y}$

  Notar que si y=1 equivale a $\hat{y}$ y si es 0 a $1-\hat{y}$

- Tomando logaritmo a ambos lados queda

$$\log p(y|x) = \log \hat{y}^y(1-\hat{y})^{1-y}$$

$$\log p(y|x) = y \log \hat{y} + (1-y)\log(1-\hat{y})$$

- Minimizamos lo anterior multiplicado por (-1)

  $L_{CE} = -(y \log \hat{y} + (1-y)\log(1-\hat{y}))$

Entropía cruzada

Minimizamos
$L_{CE} = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$

$$H(p, q) = -\sum_{i=1}^{n} p(x_i) \log_2 q(x_i).$$

Notar la correspondencia entre 'p' y 'y' y entre 'q' y $\hat{y}$. La fórmula de entropía cruzada se instancia de este modo.

Minimizando la entropía cruzada

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \frac{1}{m} \sum_{i=1}^{m} L_{\mathrm{CE}}(f(x^{(i)};\theta), y^{(i)})$$

- $\hat{\Theta}$ es el conjunto de parámetros, $w_1, w_2, ..., w_n$ y b
- Los supraíndices afectando a los x e y corresponden a los m casos del conjunto de entrenamiento

Agenda

- Atributos o variables de entrada
- Función de clasificación
- ➔ Método de entrenamiento
- Métodos de evaluación

Métodos de entrenamiento y evaluación

- Vimos solamente cuál es la función de pérdida y qué es lo que se optimiza al entrenar.
- El entrenamiento se hace por descenso de gradiente, se verá con Redes Neuronales. También se verán estrategias para evitar sobreajustes.
- Los métodos de evaluación se presentan por separado.