Transformación Logística

* Si p es la probabilidad de éxito de un evento, se define el ratio odds como la proporción de éxito en relación al fracaso: p/ 1 – p
* Si la probabilidad de éxito posee el valor de 0.5
  + El ratio odds es 0.5/(1-0.5)=0.5/0.5=1
  + Se lee un éxito por un fracaso
* Si la probabilidad de éxito posee el valor de 0.75
  + El ratio odds es 0.75/(1-0.75)=0.75/0.25=3
  + Se lee tres éxitos por un fracaso
* Si la probabilidad de éxito posee el valor de 0.25
  + El ratio odds es 0.25/(1-0.25)=0.25/0.75=1/3=0.33
  + Se lee 0.33 éxitos por un fracaso
  + O un éxito cada 3 fracasos
* Mientras mayor sea la probabilidad de éxito mayor será el ratio odds

|  |  |
| --- | --- |
| P | Ratio odds |
| 0.1 | 0.11 |
| 0.3 | 0.43 |
| 0.5 | 1.0 |
| 0.7 | 2.3 |
| 0.9 | 9.0 |
| 0.99 | 99.0 |
| 0.999 | 999.0 |

* La función logit se define como el logaritmo natural del radio odds
  + Logit(p)=log(p/1-p)=log(p)-log(1-p)

|  |  |  |
| --- | --- | --- |
| P | Ratio odds | logit |
| 0.1 | 0.11 | -0.95 |
| 0.3 | 0.43 | -0.37 |
| 0.5 | 1.0 | 0.0 |
| 0.7 | 2.3 | 0.37 |
| 0.9 | 9.0 | 0.95 |
| 0.99 | 99.0 | 1.99 |
| 0.999 | 999.0 | 2.99 |

\*Los valores de la función logit van desde [-infinito,+infinito]

**Transformación Logística**

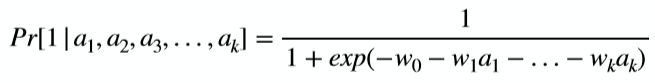
* La función logística se define como P(t)=1/1+e^(-t)
* Esta función se utiliza para estimar la probabilidad
* En la regresión logística, la salida en sí misma es tomada como el resultado esperado.
  + Sirve para predecir valores
* En **problemas de clasificación**, la salida sirve como frontera de decisión.
  + Sirve para **determinar la pertenencia** a determinada clase
* La **regresión logística** permite determinar la **probabilidad** de que una instancia pertenezca a una clase.
* Suponiendo que solo se tienen 2 clases para clasificar.
* La probabilidad que una instancia pertenezca a determinada clase se puede describir como:

Pr[1|a1,a2,a3,...,ak]

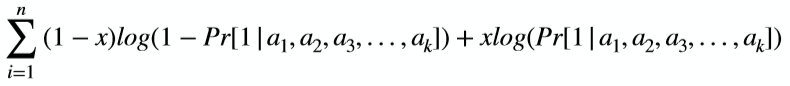
* La regresión logística reemplaza la variable objetiva original usando la función lineal siguiente:

log( Pr[1|a1,a2,a3,...,ak]/(1−Pr[1|a1,a2,a3,...,ak]))

* Los valores resultantes ya no están restringidos al intervalo de 0 a 1, puede estar en cualquier lugar entre [-infinito,+infinito]
* Para calcular la probabilidad del éxito se usa la función logística
* De esta manera:



* De forma similar a la regresión lineal, los pesos deben ser encontrados a partir de los datos de entrenamiento
  + En la regresión lineal el ajuste se hace minimizando el cuadrado de las diferencias del valor real y el valor predicho.
  + En la regresión logística se usa el logaritmo de la probabilidad en su lugar
  + De estos se obtiene:



* + - Donde x puede ser 0 o 1
    - Se busca maximizar la sumatoria.