

Aprendizaje Automático

Departamento de Informática – UC3M

TUTORIAL 8 – Logistic Regression

2023

1

Tutorial 8

Recordando teoría. Logistic Regression

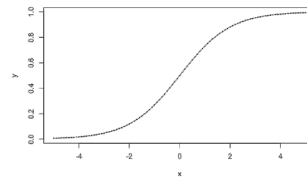
- La función logística (o sigmoide) es la aplicación particular de la función lineal de los datos a un funcional

$$y(\mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + w_0)$$

- Donde la sigmoide se define como:

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

- Es una función suave y diferenciable



2

Tutorial 8

Recordando teoría. Clasificación Logística

- Dado que está acotada $[0,1]$, se puede usar para modelar la probabilidad de una clase:

$$p(C = 0|\mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + w_0) = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + w_0)}}$$

- Si se suponen dos clases, ¿ $p(C=1|\mathbf{x})$?
- Sabiendo que: $p(C=1|\mathbf{x}) + p(C=0|\mathbf{x}) = 1$
- Entonces:

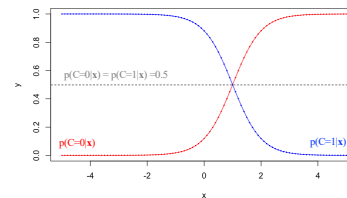
$$p(C = 1|\mathbf{x}) = 1 - \sigma(\mathbf{w}^T \mathbf{x} + w_0) = 1 - \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + w_0)}} = \frac{e^{-(\mathbf{w}^T \mathbf{x} + w_0)}}{1 + e^{-(\mathbf{w}^T \mathbf{x} + w_0)}}$$

3

Tutorial 8

Recordando teoría. Clasificación Logística

- ¿Cómo es la frontera de decisión (*decision boundary*) para la regresión logística?
- Es decir, ¿cómo se separa una clase de la otra? ¿Será una separación lineal, cuadrática, exponencial...?
 - Si se tienen dos clases con la misma probabilidad: $p(C=0|\mathbf{x}) = p(C=1|\mathbf{x}) = 0.5$



4

Tutorial 8

Clasificación Logística. Sklearn. Parámetros

- Necesita que las entradas sean numéricas y escaladas

```
LogisticRegression(random_state = 42)
```
- Parámetros relevantes:
 - **Penalty**. Regularización aplicada {None, 'l1', 'l2', 'elasticnet'}
 - **C**. Inverso de la fuerza de regularización.
 - **l1_ratio**. El parámetro de mezcla Elastic-Net
 - **Solver**. Algoritmo de optimización a usar {'lbfgs', 'liblinear', 'newton-cg', 'newton-cholesky', 'sag', 'saga'}
 - **class_weight**. Con el valor 'balanced' fuerza a que en el proceso de aprendizaje se tomen las clases de modo balanceado.

5

Tutorial 8

Selección de características por filtrado

- Varianza (VarianceThreshold)
- Mutual information. Ganancia de información (mutual_info_classif)
- Test F de ANOVA (f_classif)
- Chi2. El test chi2 parte de la hipótesis nula de que dos variables son independientes (SelectKBest)

6