KSCHOOL

Del Dato al Conocimiento

Métodos de Regularización y Estadística Bayesiana

Antonio Pita Lozano

Máster en Data Science

Métodos de Regularización

Del Dato
al Conocimiento

Los modelos de regresión y clasificación pueden presentar problemas en su construcción como sobreajuste o multicolinealidad. Los métodos de regularización son técnicas que permiten mejorar la capacidad de generalización de los modelos.

Principales técnicas:

- Regularización Ridge
- Regularización Lasso
- Regularización Elastic Net

Métodos de Regularización: Ridge

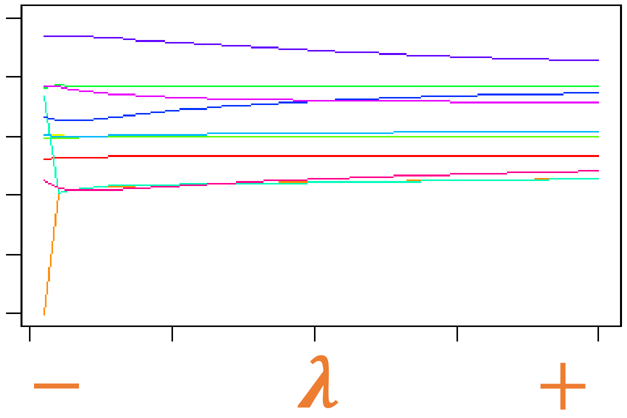
La Regularización Ridge pretende suavizar los coeficientes para mitigar las oscilaciones provocadas por la sobreestimación o la multicolinealidad fuerte.

Función de coste $g(\beta_0, \beta_1, \dots, \beta_m) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a - \beta_1 x_{i1} - \dots - \beta_m x_{im})^2$

Función de coste con Regularización $g'(\beta_0, \beta_1, \dots, \beta_m, \lambda) = \sum_{i=1}^{n} \varepsilon_i^2 + \lambda \|\beta\|_{L2}^2 =$

$$\sum_{i=1}^{n} (y_i - a - \beta_1 x_{i1} - \dots - \beta_m x_{im})^2 + \lambda(\beta_1^2 + \dots + \beta_m^2)$$

Se reduce el valor de los coeficientes



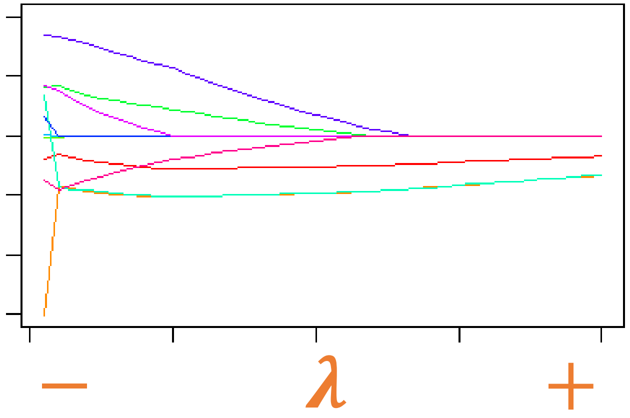
La Regularización Lasso pretende simplificar el modelo entrenado mediante la reducción de variables.

Función de coste $g(\beta_0, \beta_1, \dots, \beta_m) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n}(y_i - a - \beta_1 x_{i1} - \cdots - \beta_m x_{im})^2$

Función de coste con Regularización $g'(\beta_0, \beta_1, \dots, \beta_m, \lambda) = \sum_{i=1}^{n} \varepsilon_i^2 + \lambda \|\beta\|_{L1} =$

$$\sum_{i=1}^{n}(y_i - a - \beta_1 x_{i1} - \cdots - \beta_m x_{im})^2 + \lambda(|\beta_1| + \cdots + |\beta_m|)$$

Se reduce el número de variables



La Regularización Elastic Net es una generalización de las regularizaciones Ridge y Lasso.

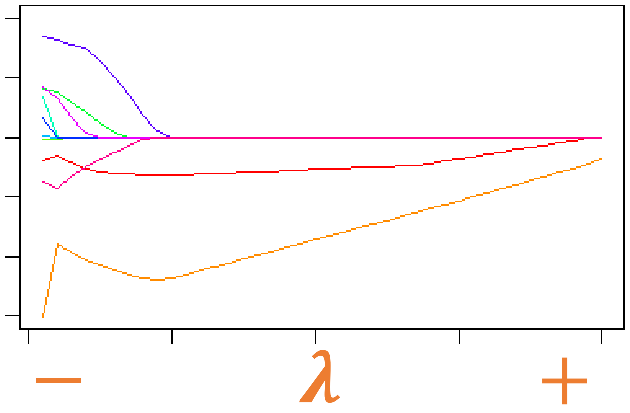
Función de coste $g(\beta_0, \beta_1, \ldots, \beta_m) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n}(y_i - a - \beta_1 x_{i1} - \cdots - \beta_m x_{im})^2$

Función de coste con Regularización

$$g'(\beta_0, \beta_1, \ldots, \beta_m, \lambda) = \sum_{i=1}^{n} \varepsilon_i^2 + \lambda\left(\alpha\|\beta\|_{L1} + (1-\alpha)\|\beta\|_{L2}^2\right) =$$

$$\sum_{i=1}^{n}(y_i - a - \beta_1 x_{i1} - \cdots - \beta_m x_{im})^2 + \lambda * (\alpha(|\beta_1| + \cdots + |\beta_m|) + (1-\alpha)(\beta_1^2 + \cdots + \beta_m^2))$$

$$\alpha = 0 \rightarrow Ridge$$
$$\alpha = 1 \rightarrow Lasso$$



Métodos de Regularización

La **Estadística Bayesiana** basa sus teorías en que la probabilidad se va modificando en función de las evidencias conocidas, así, la probabilidad de ocurrencia de un evento A (P(A) llamada **prior**) se modifica al conocer una evidencia B. La nueva probabilidad se llama **posterior** y se denota por P(A|B)).

Teorema de Bayes: siendo A y B dos eventos

$$\text{posterior} \quad P(A|B) = \frac{\overset{\text{prior}}{P(A)} * P(B|A) \ \text{likelihood}}{P(B) \ \text{evidencie}}$$

Teorema de la Probabilidad Total: Siendo $A_1, \dots, A_n$ una partición:

$$P(A) = P(A_1) * P(A|A_1) + \cdots + P(A_n) * P(A|A_n)$$

Naive Bayes Classifier

Del Dato al Conocimiento

Objetivo

Establecer una estimación de la probabilidad de ocurrencia de un evento (o grupo de eventos) en función de las evidencias conocidas.

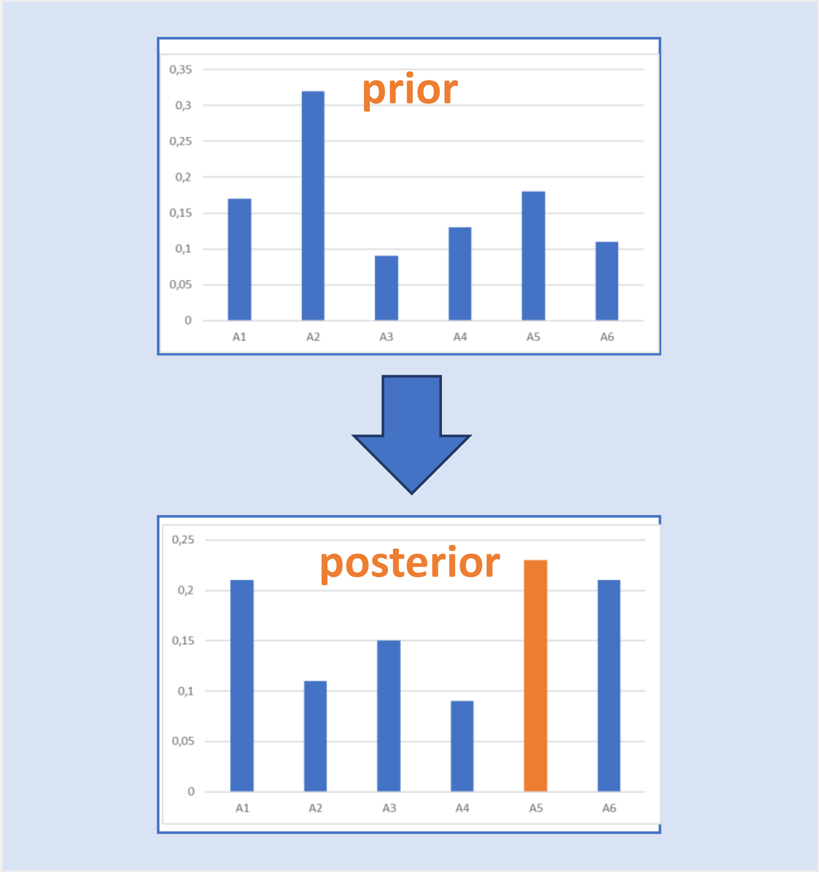
Desarrollo

Para la estimación de la probabilidad, es necesario calcular la probabilidad del evento condicionada a la evidencia, llamada posterior. Para ello se utilizan el teorema de Bayes, el teorema de la probabilidad total y la regla de la cadena.

$$\hat{y} = \arg\max_{k\in\{1,\ldots,K\}} p(A_k|x_1,\ldots,x_n) = \arg\max_{k\in\{1,\ldots,K\}} p(A_k)\prod_{i=1}^{n} p(x_i|A_k)$$

El algoritmo considera de forma ingenua (naive) que las evidencias son condicionalmente independientes entre si para que los cálculos sean sencillos y computacionalmente accesibles.

El algoritmo permite clasificar entre varias categorías ($A_1, \ldots, A_k$).



KSCHOOL

Del Dato al Conocimiento

el rincón de Antonio Pita

https://antoniopita.blog

Del Dato
al Conocimiento

Métodos de Regularización y Estadística Bayesiana

Antonio Pita Lozano

Máster en Data Science

https://www.linkedin.com/in/antoniopitalozano/

@anto_pita