

Guía de estudio - Modelos de ensamble (Parte III)



¡Hola! Te damos la bienvenida a esta nueva guía de estudio.

En esta guía abordaremos dos problemas que sufren los modelos de aprendizaje automático afectando su precisión y capacidad de generalización: el sub aprendizaje (underfitting) y al sobreaprendizaje (overfitting). El primero se manifiesta cuando un modelo es demasiado “simple” para capturar la complejidad inherente al fenómeno por medio de los datos con los que se entrena, lo que resulta en predicciones inexactas y poco confiables. Por otro lado, el sobreaprendizaje se produce cuando un modelo se ajusta demasiado a los datos de entrenamiento, perdiendo la capacidad de generalizar a nuevas observaciones y provocando un rendimiento deficiente frente a estas muestras que no ha visto el modelo.

Para resolver estos problemas se utilizan técnicas de regularización, como la Norma L1 Lasso, Norma L2 Ridge y Elastic Net (que ya abordamos inicialmente en el módulo de Machine Learning). La normalización L1 introduce una penalización en la magnitud absoluta de los coeficientes del modelo. Esta técnica favorece la selección de características relevantes al forzar algunos coeficientes a cero, lo que simplifica el modelo y previene el sobreaprendizaje por la eliminación de características menos importantes.

Por su parte, la regularización Ridge penaliza la magnitud al cuadrado de los coeficientes del modelo. Esto ayuda a evitar el sobreaprendizaje disminuyendo la influencia de los atributos con menor relevancia sin descartarlos completamente, y manteniendo así una mayor estabilidad en el modelo.

Elastic Net combina las penalizaciones de L1 y L2, aprovechando las ventajas de ambos enfoques. Al combinar estas dos regularizaciones, Elastic Net logra un equilibrio entre la selección de características que nos permite Lasso y la estabilidad de Ridge, ofreciendo mayor flexibilidad y control sobre la complejidad del modelo.

Las estrategias de regularización Lasso, Ridge y Elastic Net ofrecen enfoques poderosos para abordar los desafíos de subaprendizaje y sobreaprendizaje por medio del ajuste de la complejidad del modelo. La elección de la estrategia a utilizar depende de la naturaleza del problema y en que estemos interesados ya sea la selección de atributos, la magnitud de los coeficientes o ambos.

A lo largo de la guía daremos a conocer en forma teórica los métodos de regularización clásicos, junto con su implementación en sklearn revisando su aplicación con datos reales y midiendo el impacto de estos en los resultados.

¡Vamos con todo!



Tabla de contenidos

Guía de estudio - Modelos de ensamble (parte III)	1
Tabla de contenidos	2
Sub aprendizaje y Sobreaprendizaje	3
Sub aprendizaje	3
Sobreaprendizaje	3
Importancia de regularizar	4
Regularización Ridge L2	5
Regularización Ridge	6
Actividad guiada: Penalización con norma Ridge	6
Regularización Lasso L1	6
Actividad guiada: Implementación de regularización Lasso	7
Regularización Elastic Net	13
Implementación regularización Elastic Net	13
Referencias bibliográficas	16



¡Comencemos!

Sub aprendizaje y Sobreaprendizaje

El sub aprendizaje y el sobreaprendizaje son dos problemas fundamentales que afectan al aprendizaje automático en cuanto a su precisión y su capacidad de generalización.

Sub aprendizaje

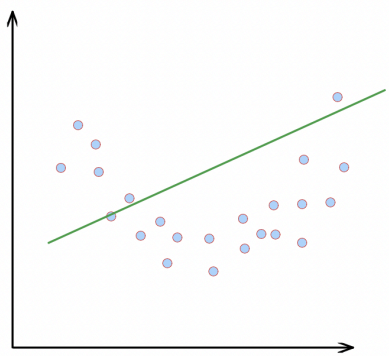


Figura 1. Modelo con problema de sub-ajuste
Fuente: DesafíoLatam

En la Figura 1, se muestra un modelo (la recta) en el que la simpleza de un recta no aproxima correctamente a la complejidad de los datos, lo que resulta en predicciones con baja precisión esto tanto en los datos de entrenamiento como en los de prueba. Para solucionar este problema podemos recolectar más datos para el entrenamiento (data augmentation).

Sobreaprendizaje

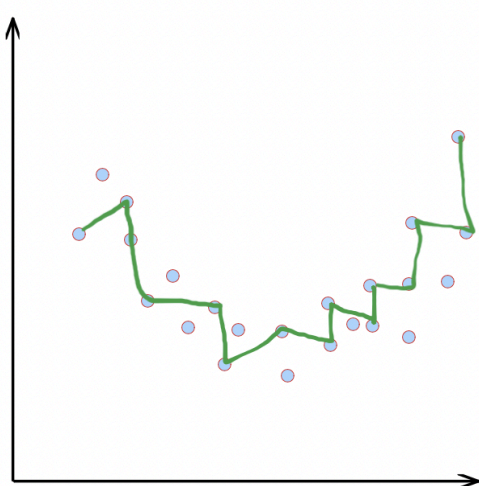


Figura 2. Modelo con problema de sobreajuste
Fuente: DesafíoLatam

El sobreaprendizaje se genera cuando el modelo se ajusta demasiado a los datos de entrenamiento, aprendiendo el ruido y perdiendo la capacidad de generalización para datos

desconocidos, esto se aprecia en la Figura 2. Para mejorar estos problemas aplicamos estrategias de regularización.

Un modelo saludable que no presenta problemas ni de sub aprendizaje ni sobreaprendizaje se muestra en la Figura 3.

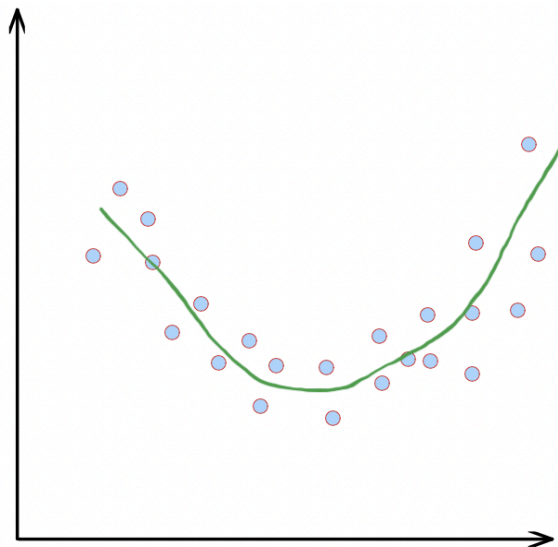


Figura 3. Modelo bajo sesgo y baja varianza
Fuente: Desafío Latam

Importancia de regularizar

Además de prevenir el sobreajuste, la regularización nos permite evitar problemas de multicolinealidad. Cuando los modelos usan variables regresoras que resultan estar altamente correlacionadas y esto provoca problemas de estabilidad en ciertos modelos, la Norma L2 ayuda a disminuir la multicolinealidad.

Además, nos permite realizar una selección automática de características: La regularización Lasso L1 como Elastic Net permiten seleccionar aquellos atributos importantes, descartando aquellos de menor relevancia en el ajuste del modelo.

Regularización Ridge L2

La regularización Ridge penaliza aquellos parámetros estimados con valores grandes y lo hace por medio de la Norma L2.

$$L2 : ||\mathbf{w}\|_2^2 = \sum_{j=1}^m w_j^2$$

En el caso de una regresión lineal, para una función de pérdida de el error cuadrático medio tenemos que:

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^p \beta_i^2$$

Cuando nuestro modelo otorgue mucha importancia a una variable regresora por medio de un valor más alto en el coeficiente asociado, se agrega una penalización alta (ya que también es cuadrática) induciendo que se permita un incremento de algún coeficiente cuando se genere una disminución proporcional en la suma de los cuadrados. Lo que estaremos haciendo con esto es agregando sesgo al modelo y empujando la elección de un modelo más simple.

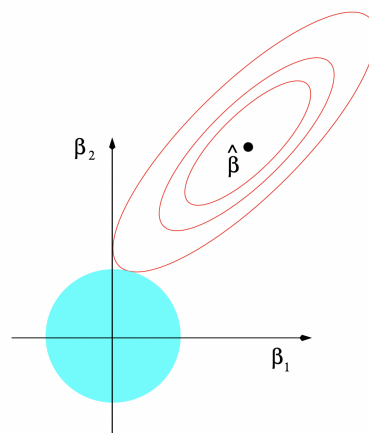


Figura 4. Representación geométrica de regularización normal L2

Fuente: Trevor Hastie, Robert Tibshirani, Martin Wainwright. SLS Statistical Learning with Sparsity The Lasso and Generalizations.

En la Figura 4, vemos una representación geométrica de la norma L2 que se muestra como una esfera, y las curvas de nivel como elipses centradas en el estimador OLS para beta. El objetivo es minimizar la suma entre OLS y la penalización. De esta manera, nos ubicamos en la superficie de la esfera para hacer coincidir con alguna curva de nivel de OLS, lo que significa que los coeficientes estimados no podrán ser iguales a cero.

El hiper parámetro lambda que acompaña al penalizador corresponde al grado de penalización que se desea incorporar al modelo, es decir, cuando lambda es cero estaremos en presencia de un modelo sin penalización. Debemos entregar un valor fijo inicial a este hiper parámetro, valor que puede ayudar o no a nuestro modelo por lo cual debe ser sintonizado, es decir, debemos aplicar estrategias para optimizarlo.



Actividad guiada: Penalización con norma Ridge

Analiza lo propuesto en el archivo **01 - Implementación de normas**, Sección “Penalización con norma Ridge”

Regularización Lasso L1

Para la regularización Lasso se usa la Norma L1, que penaliza los parámetros estimados con el valor absoluto de cada coeficiente.

$$L1 : ||\mathbf{w}||_1 = \sum_{j=1}^m |w_j|$$

En el caso de una regresión lineal, para una función de pérdida de el error cuadrático medio tenemos:

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^p |\beta_i|$$

Si analizamos esta nueva penalización en forma gráfica veremos esquinas para el penalizador, como se muestra en la Figura 5 en que el rombo es el penalizador y las curvas de nivel la función OLS centrado en beta. Debemos buscar la minimización aditiva de ambos componentes, lo que permite que los coeficientes puedan ser cero. Esta característica de Lasso nos permite seleccionar o descartar aquellas características que tengan menor importancia para la función de pérdida aplicada (para este ejemplo, OLS) por lo que ahora podemos pensar en Lasso para ser usado inicialmente en problemas de reducción de dimensionalidad.

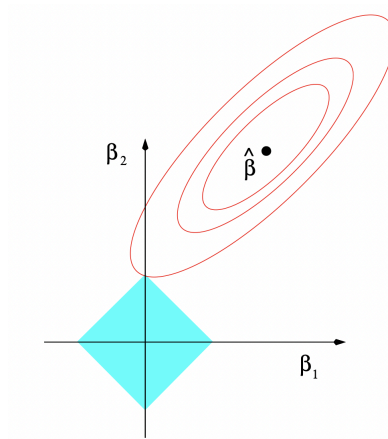


Figura 5. Representación geométrica de regularización normal L1

Fuente: Trevor Hastie, Robert Tibshirani, Martin Wainwright. SLS Statistical Learning with Sparsity The Lasso and Generalizations.

Al igual que en Ridge, en Lasso tenemos un hiper parámetro lambda que nos permite controlar el grado de penalización que se aplica al modelo, el cual debe ser sintonizado adecuadamente.



Actividad guiada: Implementación de regularización Lasso

Aplicaremos la regularización Lasso para el mismo conjunto de datos que en Ridge (abalone). En sklearn utilizaremos las librerías “Lasso” y “LassoCV”, que implementan validación cruzada. Analiza para esto el archivo de Jupyter anterior, sección “Regularización Lasso L1”

Regularización Elastic Net

La idea fundamental de Elastic Net es conseguir la combinación de penalización Ridge junto con la que ofrece Lasso que permite la selección de atributos.

$$\hat{\beta}^{enet} = \arg \min_{\beta} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda_1 \sum_{i=1}^p |\beta_i| + \lambda_2 \sum_{i=1}^p \beta_i^2$$

En este nuevo modelo de regularización necesitaremos sintonizar dos hiper parámetros lambda1 y lambda2 asociados a Norma L1 (Lasso) y Normal L2 (Ridge).



Actividad guiada: Implementación de regularización Elastic Net

Aplicaremos la regularización Lasso para el mismo conjunto de datos que en Ridge (abalone). En sklearn utilizaremos las librerías “Lasso” y “LassoCV”, que implementan validación cruzada.

En sklearn tenemos disponible ElasticNet y ElasticNetCV; este último método considera validación cruzada. Debemos tener en cuenta que la librería tiene un parámetro α y un l1_ratio y estos no funcionan independientemente: cuando l1_ratio toma un valor igual a cero entonces sólo estará usando la parte de la Norma L2 (Ridge), y cuando es 1.0 utiliza sólo la Norma L1 (Lasso). En ambos casos el hiper parámetro que acompaña a estas regularizaciones es α .

Analiza para esto el archivo de Jupyter de la Guía, sección “Regularización Elastic Net”

Reflexiona

- ¿Qué dificultades tiene la implementación de estas regularizaciones? ¿Qué aspectos deberíamos observar o verificar antes de aplicarlas?
- ¿Qué beneficios trae el uso de parámetros adecuados para implementar nuestro modelo?
- ¿Qué aspectos nuevos aprendiste respecto de las normas de penalización, considerando lo visto en el módulo de Machine Learning?



Referencias bibliográficas

The Elements of Statistical Learning

T. Hastie, R. Tibshirani, and J. Friedman. *Springer Series in Statistics Springer New York Inc., New York, NY, USA, (2001)*