ACÁMICA

TEMA DEL DÍA

Regresión Lineal

Hoy profundizaremos en el desarrollo de un modelo de regresión lineal, con foco en su aplicación a problemas no lineales.

También, veremos una técnica para evitar el sobreajuste: la regularización.

Agenda

Daily

Explicación: Regularización

Break

Hands-on training

Cierre

Daily

Daily

Sincronizando...

Bitácora

¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Challenge

¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Repaso de la bitácora
Regresión Lineal

Machine Learning

Aprendizaje Supervisado

Regresión

Modelos

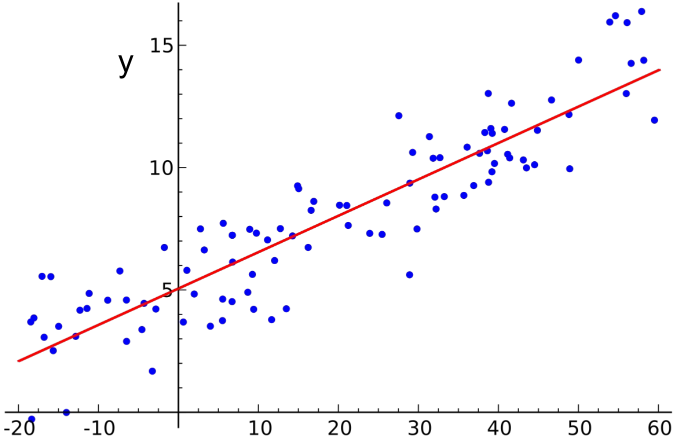
- **Regresión Lineal**
- **Árbol de Decisión**
- **k-nearest neighbors**
- etc…

Problema de regresión

Consiste en predecir una respuesta numérica Y en base a atributos X_1, X_2, ..., X_p.

$$Y \approx f\,(X_1, X_2, \ldots, X_p)$$

El caso más sencillo
es una regresión lineal.



Problema de regresión

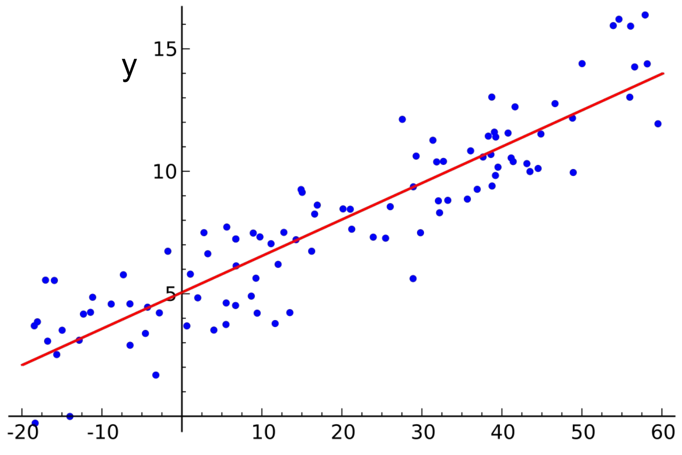
Consiste en predecir una respuesta numérica Y en base a atributos X_1, X_2, ..., X_p.

$$Y \approx f(X_1, X_2, \ldots, X_p)$$

El caso más sencillo es una regresión lineal.

Buscamos Y = mX + b que mejor ajuste a los datos:

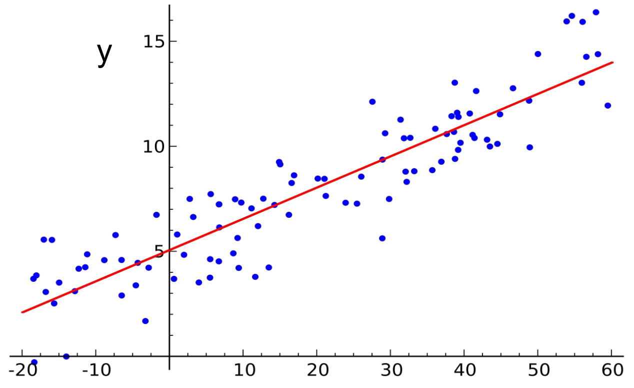
- **m: pendiente**
- **b: ordenada al origen**



Regresión lineal

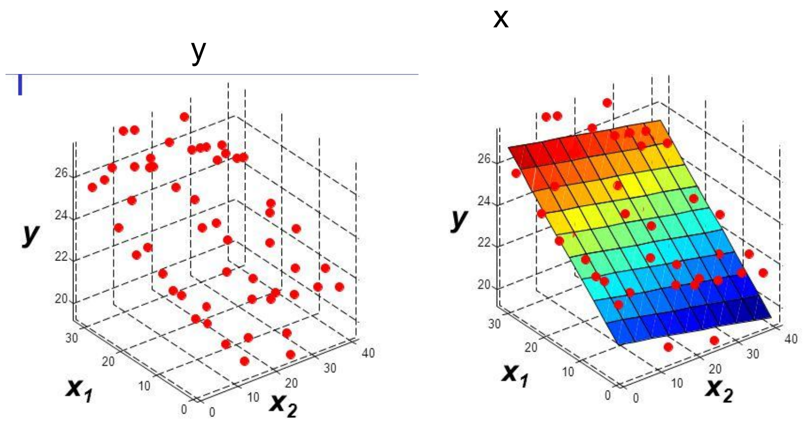
Un atributo: x_1

$Y = m_1x_1 + b$



Dos atributos: x_1, x_2

$Y = m_1x_1 + m_2x_2 + b$



Tres atributos: x_1, x_2, x_3

$Y = m_1x_1 + m_2x_2 + m_3x_3 + b$

***p* atributos: x_1, x_2, x_3, … , x_p**

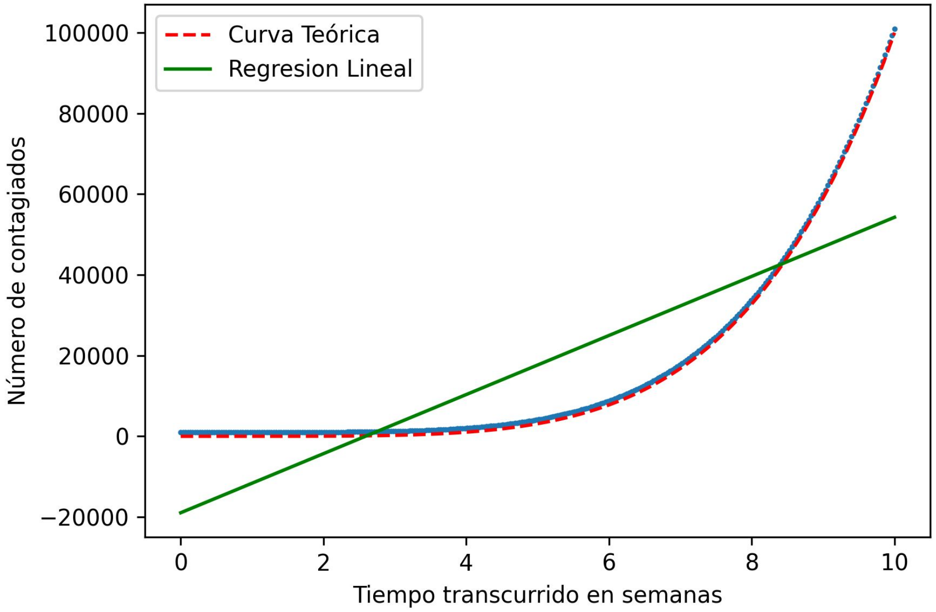
$Y = m_1x_1 + m_2x_2 + m_3x_3 + … + m_px_p + b$

Repaso de la bitácora
Adaptar el problema

Sabemos que la regresión lineal,
como su nombre lo indica,
funciona para problemas lineales.

Pero, ¿qué pasa si aplicamos una
regresión lineal a un problema no lineal?

¿Qué pasa si aplicamos una regresión lineal a un problema no lineal?



Antes de seguir, hagamos un repaso sobre qué representa cada fórmula…

Fórmulas genéricas

¡Vienen dadas por la teoría matemática!

Fórmulas funcionales (muestran la relación entre Y y X)

Son propias de nuestros datos. En los problemas que trabajamos no tenemos esta función: tratamos de estimarla.

Fórmulas genéricas

Fórmulas funcionales (muestran la relación entre Y y X)

Función lineal de 1 dimensión:

$y = ax + b \longrightarrow y = 5x + 7$

Fórmulas genéricas

Fórmulas funcionales (muestran la relación entre Y y X)

Función lineal de 1 dimensión:

$$y = ax + b \longrightarrow y = 5x + 7$$

Regresión lineal en más dimensiones:

$$y = ax_1 + bx_2 + cx_3 + \dots + d \longrightarrow y = 3x_1 + 1x_2 + 9$$

Fórmulas genéricas

Fórmulas funcionales (muestran la relación entre Y y X)

Función lineal de 1 dimensión:

$y = ax + b \longrightarrow y = 5x + 7$

Regresión lineal en más dimensiones:

$y = ax_1 + bx_2 + cx_3 + \ldots + d \longrightarrow y = 3x_1 + 1x_2 + 9$

Función de 1 dimensión **no lineal**:

$y = ax^2 + bx + c \longrightarrow y = x^2$

Al desarrollar el modelo buscamos los mejores a, b, c y d que mejor ajusten a nuestros datos. Estos se llaman **coeficientes o parámetros**.

$$y = ax_1 + bx_2 + cx_3 + \dots + d$$

El Y y los X_i van a ser tomados o generados a partir de nuestro dataset. Por eso, cada vez que agregamos un nuevo atributo implica una nueva dimensión en nuestro modelo.

Pero, ¿qué pasa si aplicamos una regresión lineal a un problema no lineal?

¿Cómo convertir un problema no lineal en uno lineal?

¿Cómo convertir un problema no lineal en uno lineal?

Si tenemos:

$$y = ax^2 + bx + c$$

¿Cómo convertir un problema no lineal en uno lineal?

Si tenemos:

De paso... ¿Cuántas dimensiones tiene?

$$y = ax^2 + bx + c$$

¿Cómo convertir un problema no lineal en uno lineal?

Si tenemos:

$$y = ax^2 + bx + c$$

¿Cómo convertir un problema no lineal en uno lineal?

Si tenemos:

$$y = ax^2 + bx + c$$

Lo puedes convertir en una relación lineal:

¿Cómo convertir un problema no lineal en uno lineal?

Si tenemos:

$$y = ax^2 + bx + c$$

Lo puedes convertir en una relación lineal:

$$y = ax_1 + bx_2 + c$$

donde $x_1 = x^2$ y $x_2 = x$

¿Cómo convertir un problema no lineal en uno lineal?

Si tenemos:

$$y = ax^2 + bx + c$$

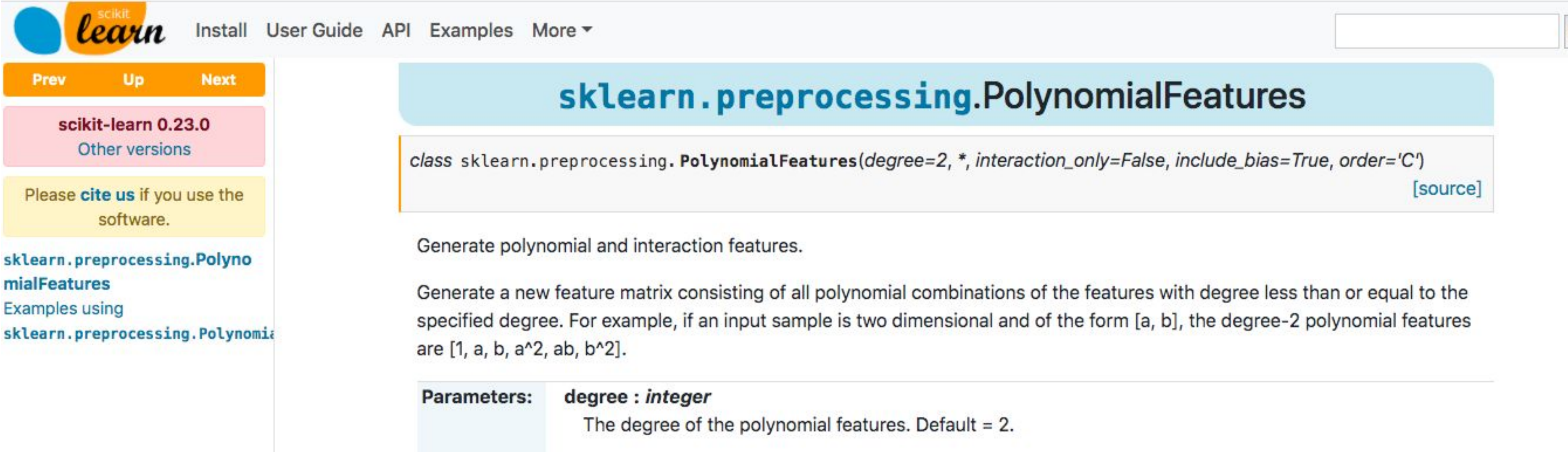
Lo puedes convertir en una relación lineal:

Nota que agregamos una nueva dimensión.

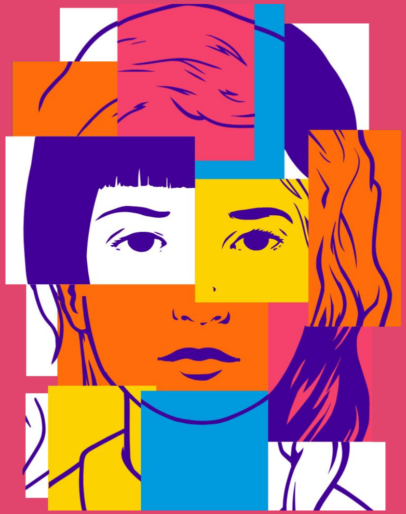
$$y = ax_1 + bx_2 + c$$

donde $x_1 = x^2$ y $x_2 = x$

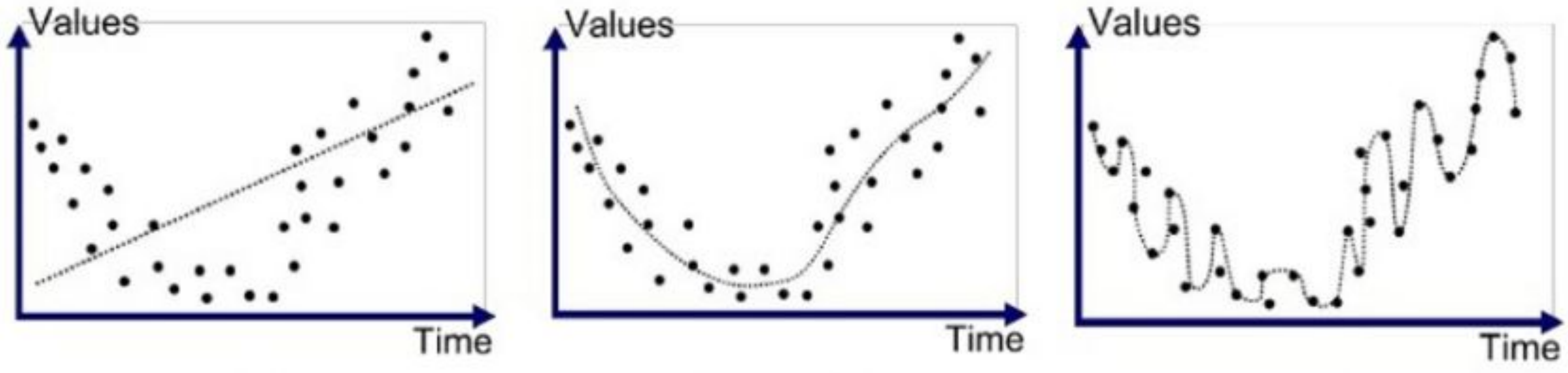
En la vida real tendremos más atributos y relaciones más complejas, por lo que usaremos Polynomial Features de Sklearn para generar nuevos atributos (¡muy buena documentación!).



Regularización



Los sospechosos de siempre…



Underfitted Good Fit/Robust Overfitted

Cómo evitarlos

Existen dos formas de evitar el “overfitting”: obtener más datos y la regularización.

- **Obtener más datos** es normalmente la mejor solución, un modelo entrenado con más datos generalizará mejor de forma natural.

- **La regularización** se realiza cuando lo anterior no es posible. Es el proceso de modular la cantidad de información que el modelo puede almacenar o añadir restricciones con la información que se permite mantener. Si el modelo solo memoriza un pequeño número de pautas, la optimización hará que el enfoque se haga en las más relevantes, mejorando la posibilidad de generalizar bien.

Cómo evitarlos

La regularización se realiza principalmente con las siguientes técnicas:

1. **Reduciendo el tamaño del modelo**: disminuyendo el número parámetros que el modelo tiene que aprender, y con ello, su capacidad de aprendizaje.

2. **Añadiendo regularización de peso**: en general, cuanto más simple es el modelo, es mejor. Mientras pueda aprender bien, un modelo más simple es menos proclive a sufrir "overfitting". Una forma común de conseguir esto es restringir la complejidad forzando sus pesos para solamente tomar pequeños valores, regularizando la distribución de valores de peso. Esto se realiza añadiendo a la función de pérdida un costo asociado a tener grandes pesos.

Objetivo

Castigar parámetros/pesos muy grandes

están asociados a overfitting.

Regularización

¿Cómo castigar parámetros/pesos muy grandes?

3 técnicas muy comunes

Regularización L1 (Ridge) y L2 (Lasso): agregan un término a la función de costo que castiga los pesos muy grandes.

Dropout: utilizado en redes neuronales; funciona como una capa que “apaga” neuronas de la capa anterior al azar

Regulación L2 o Ridge

Se agrega a la función de costo un término proporcional al cuadrado del valor de los coeficientes de peso (norma L2 de los pesos):

Parámetro de Penalización

Penalización cuadrática (L2)

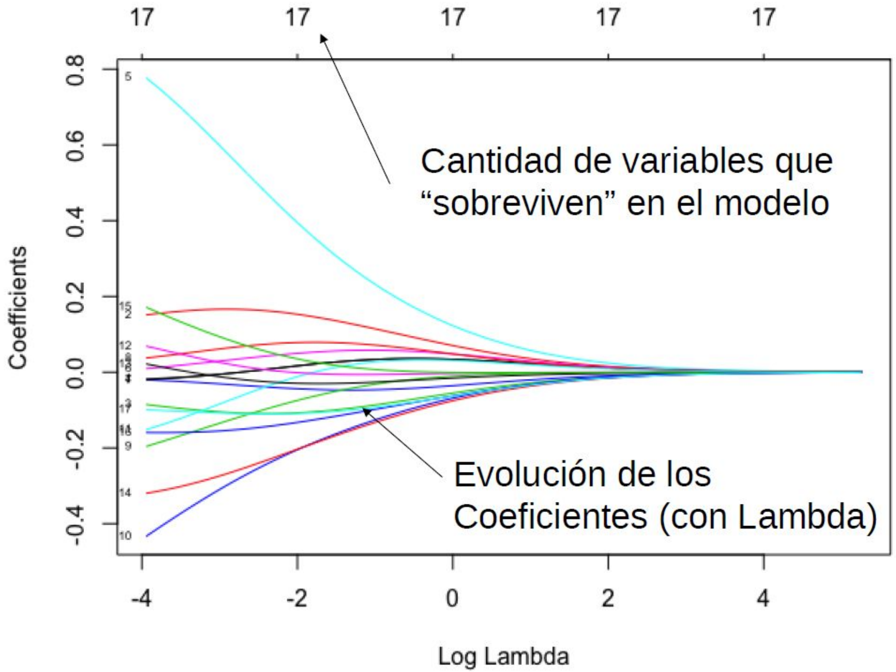
$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Falta de Ajuste

Penalización, encoge los coeficientes hacia el CERO (Shrinkage)

Regulación L2 o Ridge

Se agrega a la función de costo un término proporcional al cuadrado del valor de los coeficientes de peso (norma L2 de los pesos):



Regulación L1 o Lasso

Se agrega a la función de costo un término proporcional al valor absoluto de los coeficientes de peso (norma L1 de los pesos).

- Puede usarse como método de selección de variables

Parámetro de Penalización

Penalización lineal (L1)

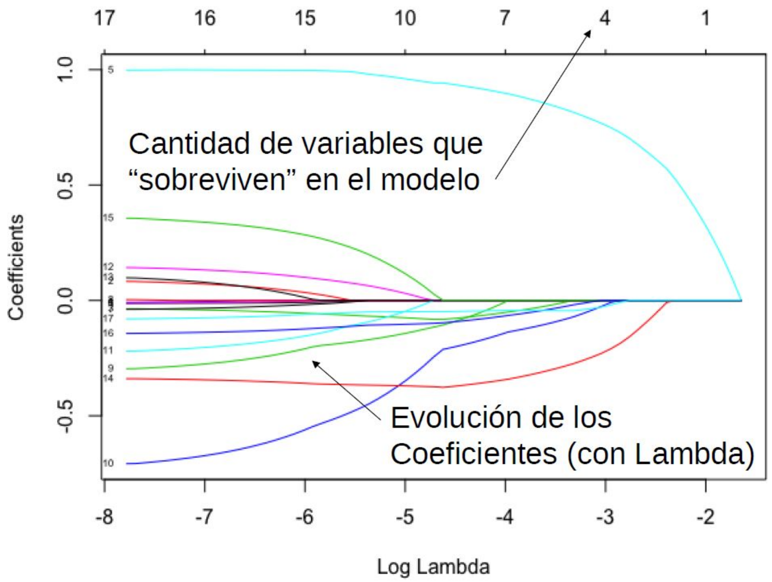
$$\boxed{\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|}$$

Falta de Ajuste

Penalización, encoge los coeficientes hacia el CERO. Convierta a algunos coeficientes en CERO

Regulación L1 o Lasso

Se agrega a la función de costo un término proporcional al valor absoluto de los coeficientes de peso (norma L1 de los pesos).



¡BREAK!

Ph. Credit: Drew Coffmann

Hands-on training

Hands-on training

DS_Bitácora_23_Regresión_Lineal.ipynb

Recursos

Recursos

- 📗 Python Data Science Handbook - Capítulo 5, *"In Depth: Linear Regression".* Lectura más que recomendada. Un poco menos técnica que la bitácora, pero con las mismas ideas. Tal vez te resulte más clara. Además, verás que no solamente existe la posibilidad de agregar atributos polinómicos, sino muchos más.
- Para una lectura integradora de estos temas, recomendamos el capítulo 9 del libro de Skiena, *The Data Science Design Manual*.
- Regularization of Linear Models with SKLearn: buen artículo en inglés que hace foco en regularización.

Para la próxima

- Termina el notebook de hoy.
- Lee la bitácora 24 y carga las dudas que tengas al Trello.

En el encuentro que viene uno/a de ustedes será seleccionado/a para mostrar cómo resolvió el challenge de la bitácora. De esta manera, ¡aprendemos todos/as de (y con) todas/as, así que vengan preparados/as.

ACÁMICA