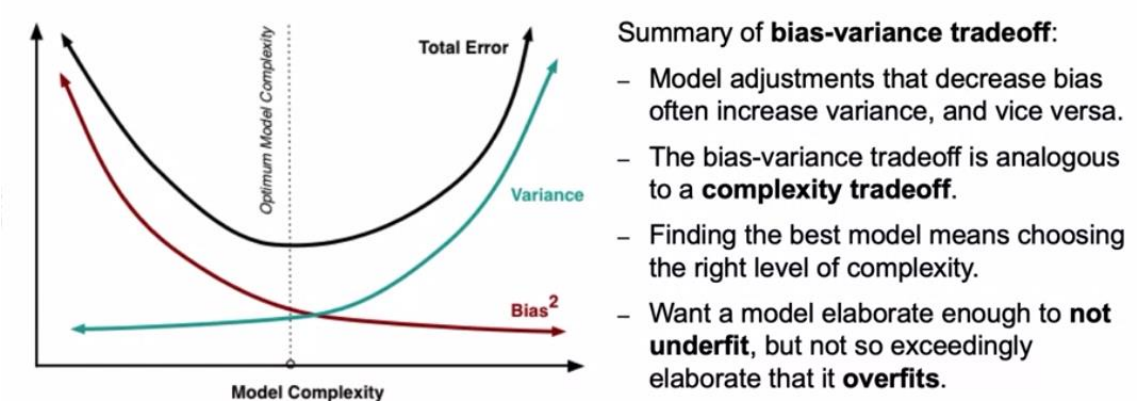
- Bias(sesgo) vs variance:

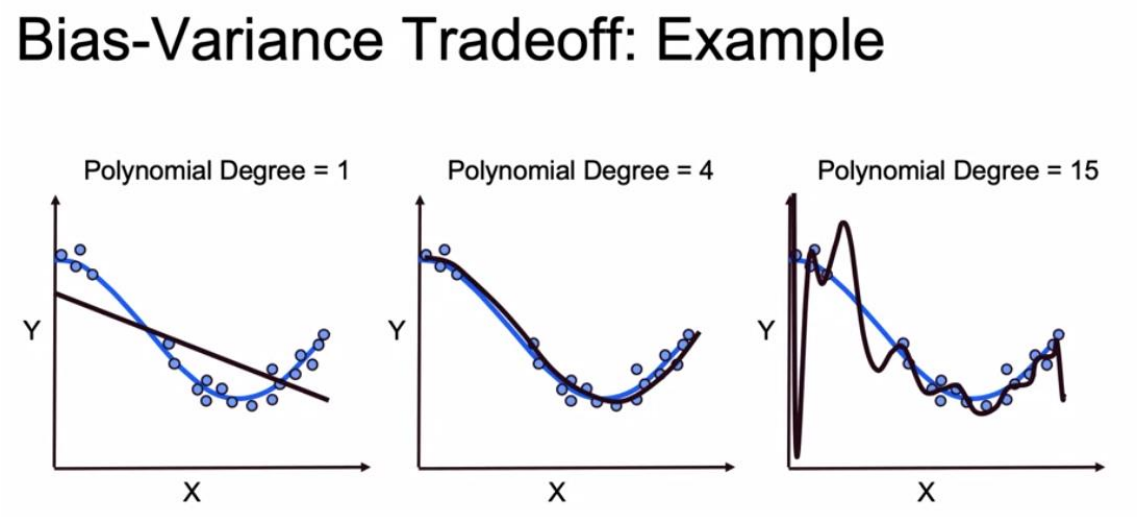
Bias = El modelo no crea relaciones consistentes entre x e y. Underfitting

Causas = Falta de información, relaciones simples

Variance = El modelo no crea relaciones consistentes entre x e y, pero no generaliza bien, por lo que funciona mal con nuevos conjuntos de datos. Overfitting

Causas = Sensibilidad a diferentes datos de entrada, overfitting



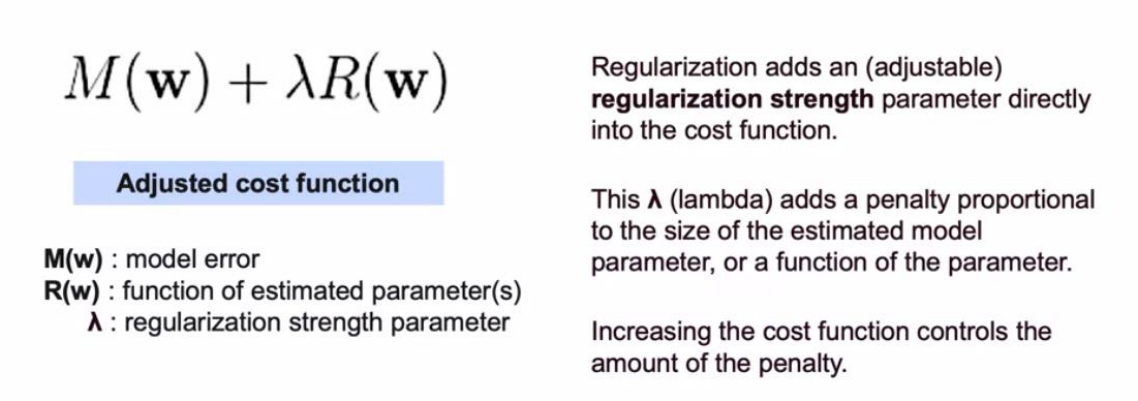


REGULARIZACIÓN

La regularización consiste en añadir un parámetro a la función de coste, lo que hará penalizar de forma extra los modelos mas complejos. Este parámetro esta representado mediante lambda.

- A mayor lambda mas posibilidad de error de Bias.
- A menor lambda mas posibilidad de error de variance (overfitting)

Es una buena técnica para prevenir el over fitting.



FEAUTURE SELECTION

Una forma de prevenir el overfitting es seleccionar las características que son realmente importantes. La regularización se encarga de ello, modificando el coeficiente de cada característica, incluso en el caso de Lasso los puede poner a cero, que sería lo mismo que eliminar dicha feature.

Ridge y Lasso explicado por copilot

Regresión Ridge (L2 Regularization)

La regresión Ridge agrega un término de penalización igual al cuadrado de la magnitud de los coeficientes multiplicado por un parámetro de regularización (λ) a la función de pérdida. Esto tiene el efecto de reducir los coeficientes, pero no los hace exactamente cero. Por lo tanto, la regresión Ridge no resulta en un modelo disperso y todos los coeficientes se incluyen en el modelo.

```
from sklearn.linear_model import Ridge
ridge = Ridge(alpha=1.0)
ridge.fit(X, y)
```

-
-
-

Regresión Lasso (L1 Regularization)

La regresión Lasso agrega un término de penalización igual al valor absoluto de la magnitud de los coeficientes multiplicado por un parámetro de regularización (λ) a la función de pérdida. Esto puede hacer que algunos de los coeficientes sean exactamente cero, lo que significa que el modelo ignora completamente algunas de las características. Por lo tanto, la regresión Lasso puede resultar en un modelo disperso y puede ser utilizada para la selección de características.

```
from sklearn.linear_model import Lasso
lasso = Lasso(alpha=1.0)
lasso.fit(X, y)
```

En ambos casos, el parámetro `alpha` es el parámetro de regularización (λ). Un valor más alto de `alpha` significa más regularización y un modelo más simple, mientras que un valor más bajo de `alpha` significa menos regularización y un modelo que se ajusta más de cerca a los datos de entrenamiento.

Por favor, ten en cuenta que estos son solo ejemplos y es posible que necesites ajustar el código para que funcione en tu entorno.

Un modelo disperso (o sparse model en inglés) en el contexto de aprendizaje automático y estadística se refiere a un modelo en el que solo un subconjunto de las características contribuye a la predicción final. En otras palabras, muchos de los coeficientes o pesos del modelo son exactamente cero. Esto significa que el modelo ignora completamente algunas de las características.

La regresión Lasso, que mencionaste anteriormente, es un ejemplo de un método que puede producir un modelo disperso. Al agregar un término de penalización basado en el valor absoluto de los coeficientes a la función de pérdida, la regresión Lasso puede hacer que algunos de los coeficientes sean exactamente cero, especialmente cuando el parámetro de regularización es grande.

Los modelos dispersos pueden ser útiles en situaciones donde tienes muchas características y sospechas que solo algunas de ellas son realmente importantes. También pueden ser útiles para mejorar la interpretabilidad del modelo, ya que es más fácil entender un modelo que solo depende de un pequeño número de características.
La elección entre la regresión Ridge y la regresión Lasso depende de tu problema y de tus datos. Aquí te dejo algunas pautas generales:

Usa la regresión Ridge si:

- Tienes muchas características y sospechas que todas o la mayoría de ellas son relevantes para tu problema. Ridge no hará que los coeficientes sean exactamente cero, por lo que no eliminará ninguna característica.

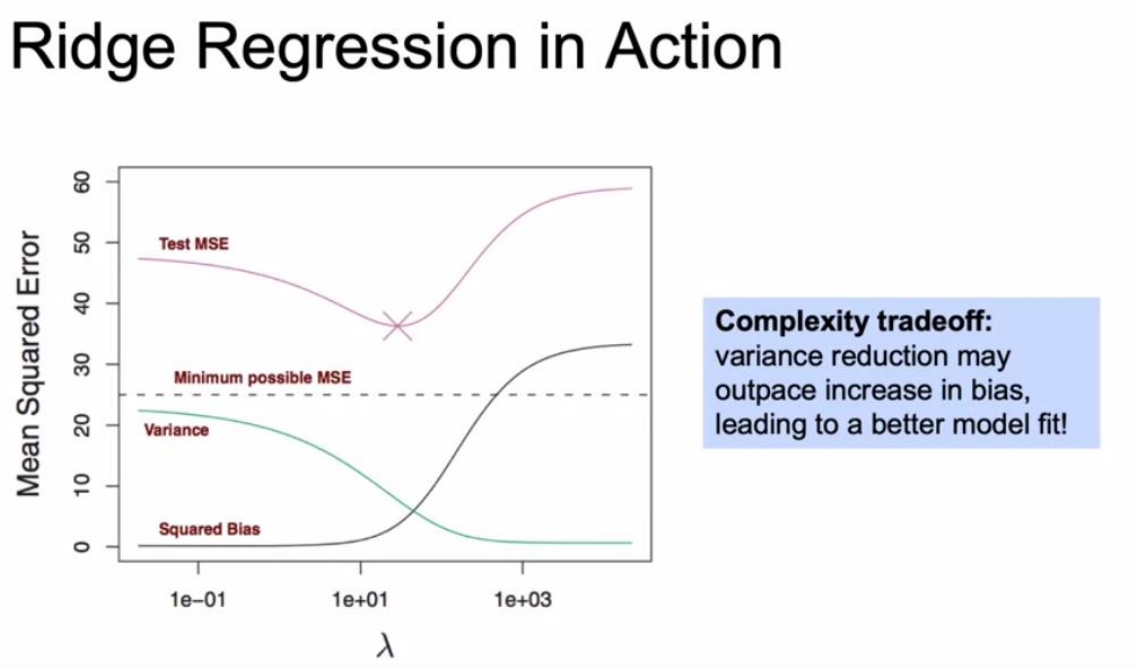
- Estás menos preocupado por hacer tu modelo interpretable. Dado que Ridge incluye todas las características, puede ser más difícil entender qué características son las más importantes.

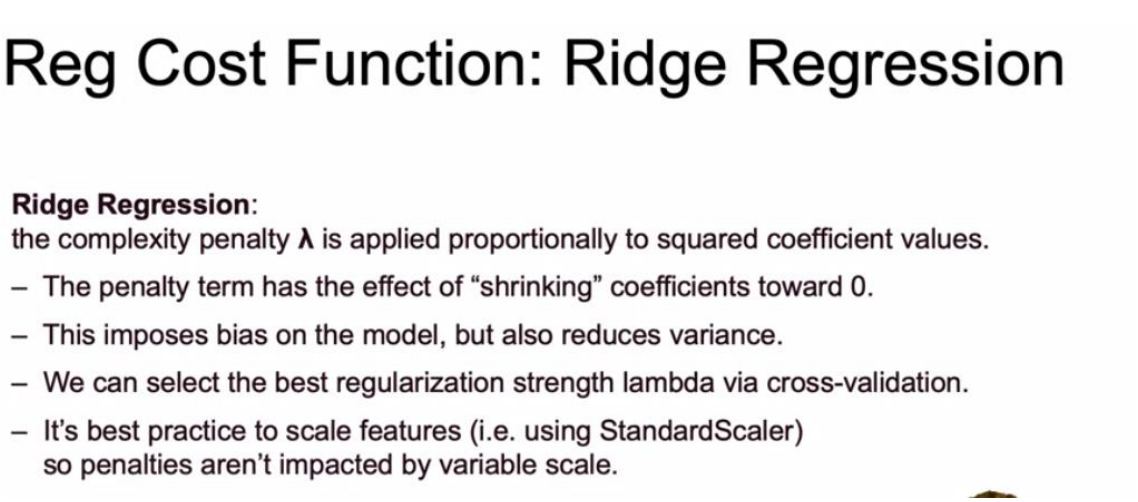
Usa la regresión Lasso si:

- Tienes muchas características y sospechas que solo algunas de ellas son relevantes. Lasso puede hacer que los coeficientes de las características irrelevantes sean exactamente cero, efectivamente eliminándolas del modelo.
- Quieres un modelo que sea fácil de interpretar. Dado que Lasso puede eliminar características, el modelo resultante puede ser más sencillo y fácil de entender.

Por último, si no estás seguro de cuál usar, podrías considerar la regresión Elastic Net, que es una combinación de Ridge y Lasso. Elastic Net incluye tanto la penalización L1 (absoluta) como la L2 (cuadrada), y tiene un parámetro que puedes ajustar para decidir cuánto peso dar a cada una. Esto puede darte lo mejor de ambos mundos.

RIDGE





A mayor coeficiente mayor penalización en la función de coste

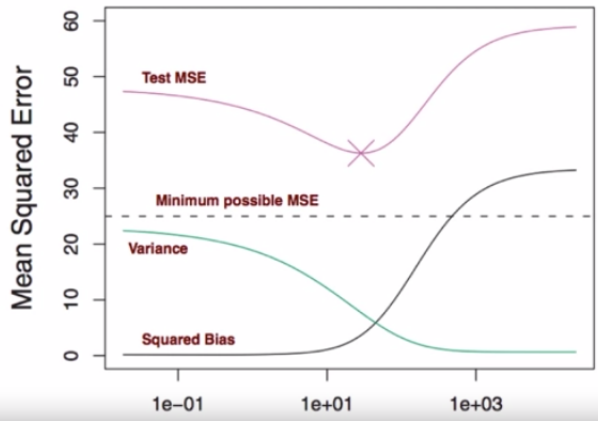
LASSO

Alternative: LASSO Regression

In **LASSO regression:** the complexity penalty **λ** (lambda) is proportional to the absolute value of coefficients.

- LASSO: Least Absolute Shrinkage and Selection Operator.
- Similar effect to **Ridge** in terms of complexity tradeoff: increasing lambda raises bias but lowers variance.
- LASSO is more likely than Ridge to perform **feature selection**, in that for a fixed **λ**, LASSO is more likely to result in coefficients being set to zero.

LASSO Regression in Action



Complexity tradeoff: variance reduction may outpace increase in bias, leading to a better model fit!

A mayor coeficiente mayor penalización en la función de coste

Elimina las características no importantes poniendo sus coeficientes a cero. Esto le da mas interpretabilidad al modelo.

Mayor coste computacional.

Elastic Net

Es un mix entre Ridge y Lasso, donde se añade un parámetro, que hará que tenga mas tendencia hacia un método u otro.

Between Ridge and LASSO: Elastic Net

$$\lambda \sum_{j=1}^{p} \left(\alpha \beta_j^2 + (1-\alpha)|\beta_j|\right)$$

Elastic Net
(hybrid approach)

Validation gives us an empirical method for selecting between different models.

LASSO's feature selection property yields an **interpretability bonus**, but may underperform if the target truly depends on many of the features.

Elastic Net, an alternative hybrid approach, introduces a new parameter ***α*** (alpha) that determines a weighted average of L1 and L2 penalties.

Elastic Net Regularization

$$J(\beta_0, \beta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\left(\beta_0 + \beta_1 x_{obs}^{(i)}\right) - y_{obs}^{(i)} \right)^2 + \boldsymbol{\lambda} \sum_{j=1}^{k} |\boldsymbol{\beta_j}| + \lambda_2 \sum_{j=1}^{k} \beta_j^2$$

Elastic Net combines penalties from both **Ridge** and **LASSO** regression.

It requires tuning of an additional parameter that determines emphasis of L1 vs. L2 regularization penalties.

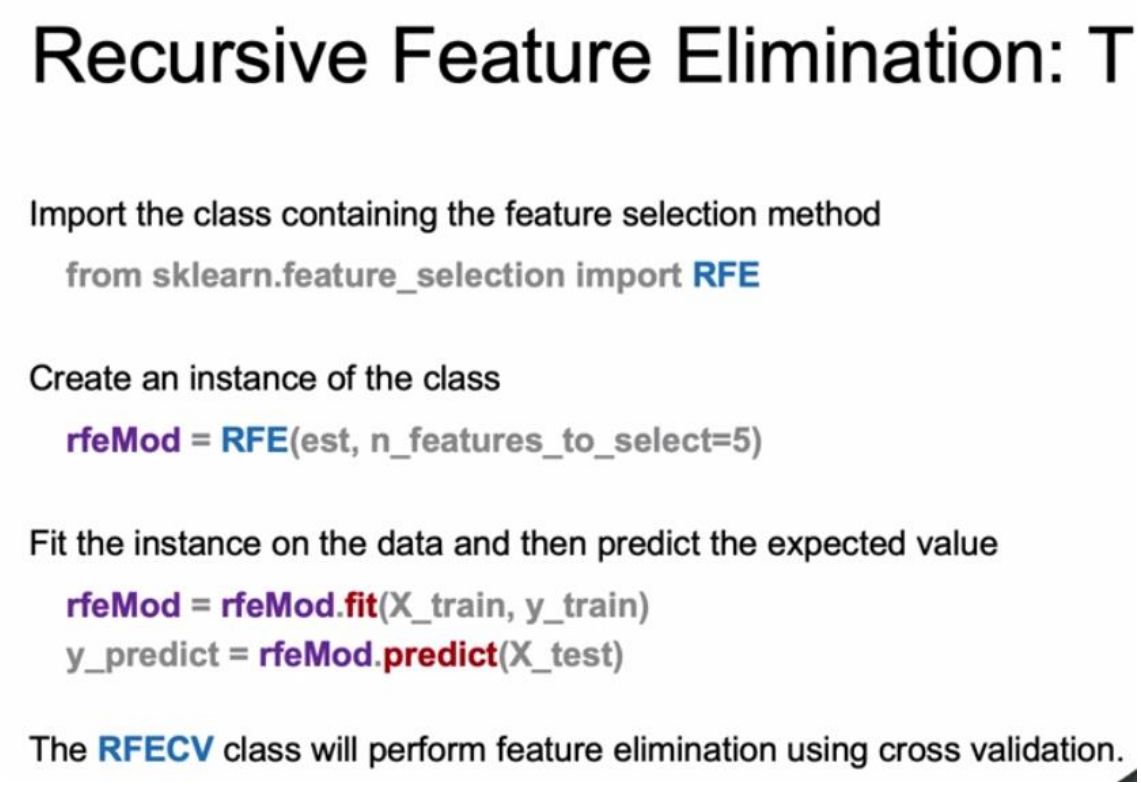
RECURSIVE FEATURE ELIMINATION

Recursive Feature Elimination

Recursive Feature Elimination (RFE) is an approach that combines:

- A model or estimation approach
- A desired number of features

RFE then repeatedly applies the model, measures feature importance, and recursively removes less important features.



Se crea una instancia, donde est es el modelo que utilizaremos y n_features_to_select, el numero total de features que conservaremos.