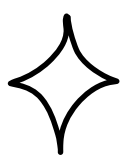
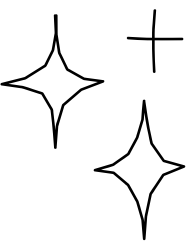
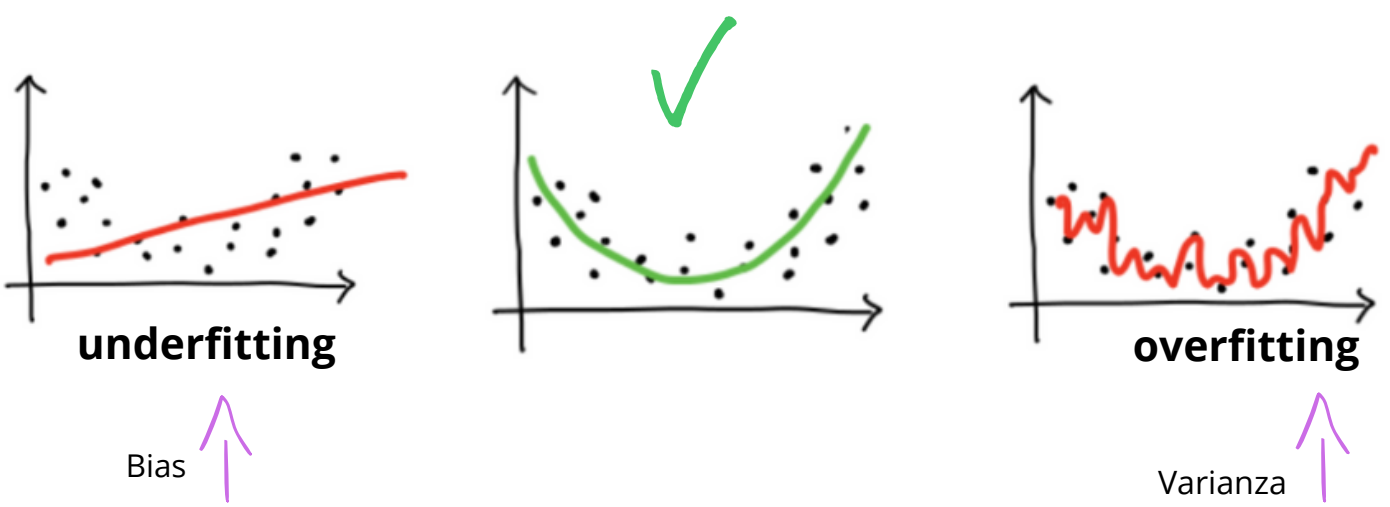


$$E(\epsilon) = \underbrace{E(\hat{Y} - Y)^2}_{\text{Sesgo}^2} + \underbrace{E(\hat{Y} - E(\hat{Y}))^2}_{\text{Varianza}} + \sigma^2$$
$$E(\epsilon) = \text{Sesgo}^2 + \text{Varianza} + \sigma^2$$

Sesgo: Es la diferencia entre la predicción de nuestro modelo y los valores verdaderos.  
Varianza: dispersion o variabilidad que pueden tener nuestras predicciones



El precio que pagamos por mejorar el sesgo es un aumento de la varianza, y viceversa.



El sesgo habitualmente presenta una relación inversa con la complejidad de los modelos. Es decir, a mayor complejidad del modelo utilizado es de esperar una menor sesgo.

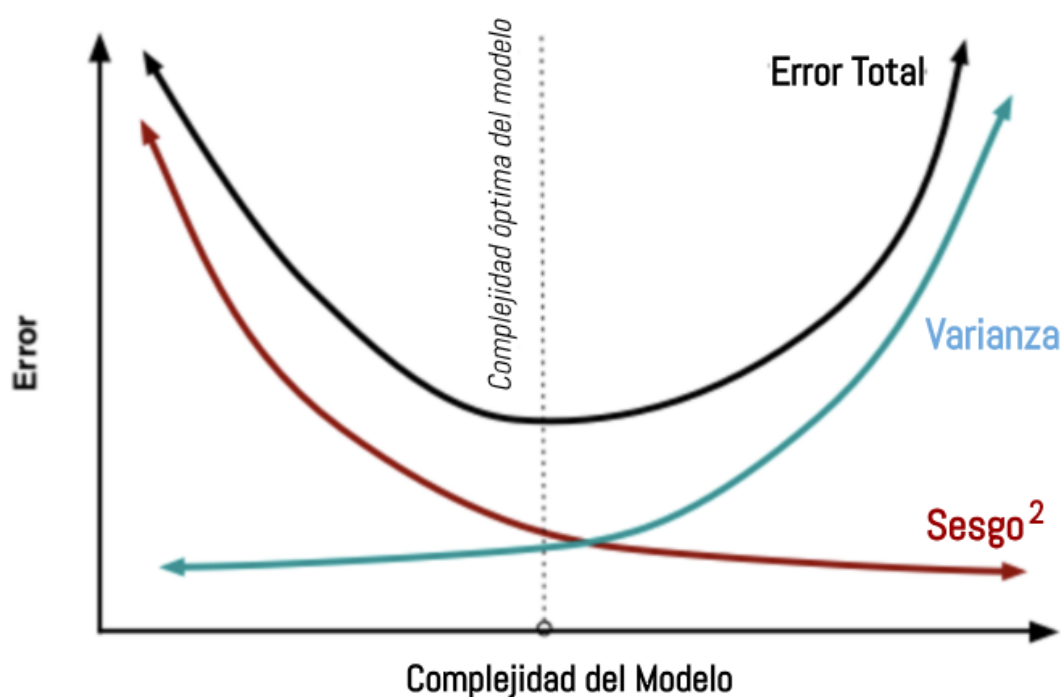
¿Qué se entiende por complejidad de un modelo? Por ejemplo, la regresión logística es un modelo más simple que un árbol de decisión, random forest es más complejo en el sentido de que utiliza un conjunto de árboles de decisión para realizar las predicciones.

La varianza también se puede relacionar con la complejidad de los modelos. A medida que aumenta la complejidad, aumentan las posibilidades de sobreajuste, es decir, la varianza aumenta. Una varianza alta significa que el modelo es demasiado complejo y presta atención en exceso a las peculiaridades del subconjunto empleado frente a los patrones generales.

Entonces, un modelo simple va a tener alto sesgo, a medida que se complejiza el modelo disminuye el sesgo pero aumenta la varianza, porque el modelo intenta aprender demasiado y puede “overfittear”.

En definitiva, lo que buscamos es un modelo que haya aprendido de los datos que se le han proporcionado (sesgo y error de entrenamiento bajos) y que sea capaz de generalizar ante nuevos datos (varianza y error de test bajos). Es decir, un modelo que no esté en régimen de overfitting ni de underfitting.

Llamamos regularización a reducir la varianza de los estimadores, logrando reducir el error del modelo a costa de resignar un poco de sesgo.



Fuente: researchgate.net