



Tecnológico de Monterrey

**Momento de Retroalimentación: Módulo 2 Análisis y Reporte sobre el desempeño del
modelo. (Portafolio Análisis)**

María Fernanda Ramírez Barragán

A01750879

ITC

Fecha: 13/09/22

Instituto Tecnológico y de Estudios Superiores de Monterrey

Inteligencia artificial avanzada para el análisis de datos

Regresión lineal múltiple

El modelo que elegí para poder trabajar con un dataset de entradas y salidas numéricas fue la regresión lineal múltiple, que es un modelo estadístico versátil para evaluar las relaciones entre un destino continuo y los predictores. La regresión lineal múltiple permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta (Y) se determina a partir de un conjunto de variables independientes llamadas predictores (X_1 , X_2 , X_3 ...).

Para entrenar a nuestro modelo dividí en 2 partes mi dataset, train y test, tomando el 80% de los datos como train y el restante 20% como test, esto porque necesitamos datos que el modelo no haya visto antes y podamos medir el accuracy y otros parámetros al comparar los datos esperados con los que el modelo predice.

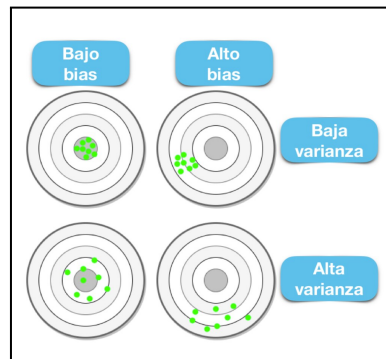
Para concluir con el análisis del modelo de regresión de sklearn y su efectividad, accuracy, etc... con nuestro dataset, calculamos con ayuda de mlxtend el sesgo promedio, la varianza promedio y la pérdida promedio esperada:

```
/-----/
/-----/
/-----/
MSE de bias_variance lib [pérdida promedio esperada]: 55.433
/-----/
Sesgo promedio: 53.651
/-----/
Varianza promedio: 1.782
/-----/
/-----/
/-----/
PS C:\Users\ferni\Documents\Tec\7ºSemestre\Models\gradient_des
```

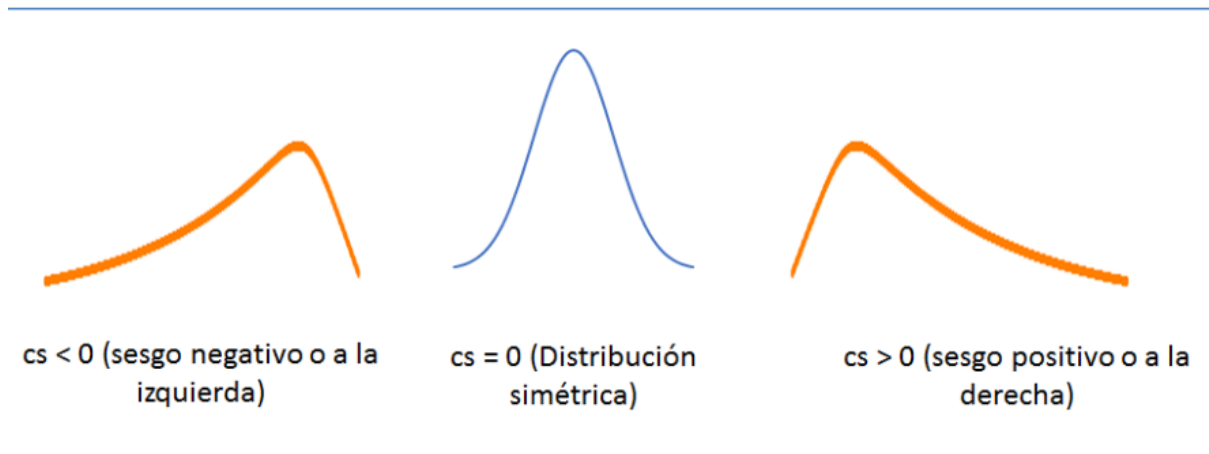
En estadística existe algo conocido como tradeoff de varianza con sesgo, es decir, dependiendo de las condiciones, el sesgo sube mientras que la varianza baja y viceversa, como podemos ver en las siguientes imágenes, hay modelos que suelen tener varianza baja y sesgo algo como también hay otros que mientras tienen las características contrarias:



Dependiendo de esto, las estimaciones se acercan o se alejan de nuestros datos, el balance ideal es que tanto la varianza como el sesgo sea bajo, de esto también depende si nuestro modelo está underfit, fit u overfit



Un coeficiente de sesgo igual a cero implica que hay una compensación entre los términos de diferencias al cubo con valores positivos y con valores negativos que contribuyen al promedio, dependiendo de nuestro sesgo el modelo tiende a inclinarse a la izquierda, al centro o a la derecha, tal y como podemos observar en la siguiente imagen:



Analizando los datos obtenidos con nuestro modelo, podemos concluir que efectivamente, dado que se trata de una regresión lineal, la varianza es baja mientras que el sesgo es algo, esto quiere decir que nuestro modelo está underfit, que nos indica que el modelo no puede predecir sobre los datos que se le proporcionan, y mucho menos sobre nuevos datos y que nuestra gráfica se ve algo así:

