



## ACTIVIDAD 2

Realizar la regresión ajustando un modelo polinómico sobre datos sintéticos. (nota docente: la guía del taller se encuentra en la sección 1 del cuaderno de jupyter llamado regresión.ipynb). Si se realizan cambios en el grado del polinomio se puede apreciar cómo estos se empiezan a sobre ajustar.

El sobreajuste es un comportamiento de aprendizaje automático no deseado que se produce cuando el modelo de aprendizaje automático proporciona predicciones precisas para los datos de entrenamiento, pero no para los datos nuevos. Cuando los científicos de datos utilizan modelos de aprendizaje automático para hacer predicciones, primero entrenan el modelo en un conjunto de datos conocido. Luego, basándose en esta información, el modelo intenta predecir los resultados para los nuevos conjuntos de datos. Un modelo sobreajustado puede proporcionar predicciones inexactas y no puede funcionar bien para todos los tipos de datos nuevos.



En este caso, el sobreajuste proviene de emplear el grado de un modelo muy grande para datos cuyo grado es menor. Es decir, el modelo es mayor que los datos en grado, lo que conduce a un sobre ajuste, (en inglés overfitting).

Ejercicio en clase: Utilizando los ejemplos de métricas de la regresión lineal, calcular las mediciones MSE, RMSE y MAE para los modelos de grado 2, 3, 6 y 20 en la regresión polinómica. ¿Qué modelo tiene el menor error? ¿Por qué se incrementa el error al incrementar el grado del polinomio?