

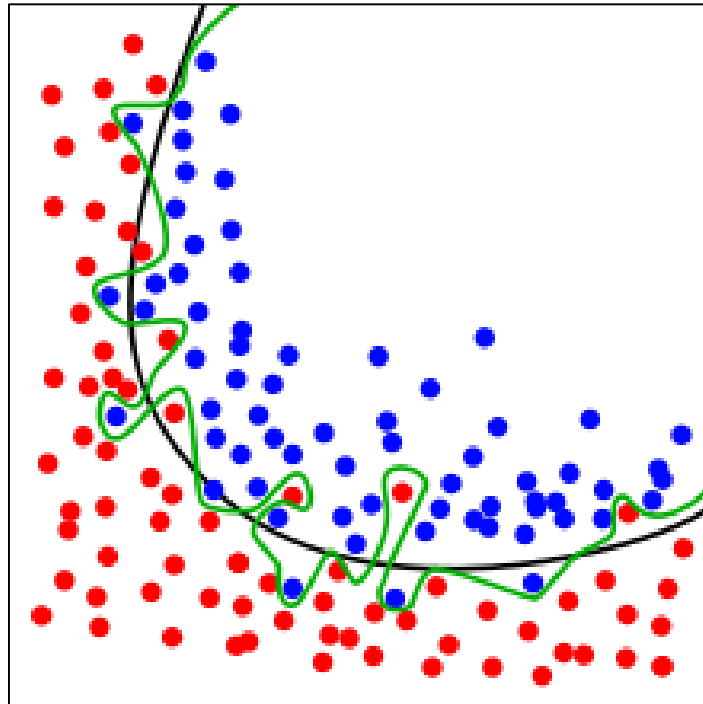
# Overfitting and Underfitting

## (Sobreajuste y Subajuste)

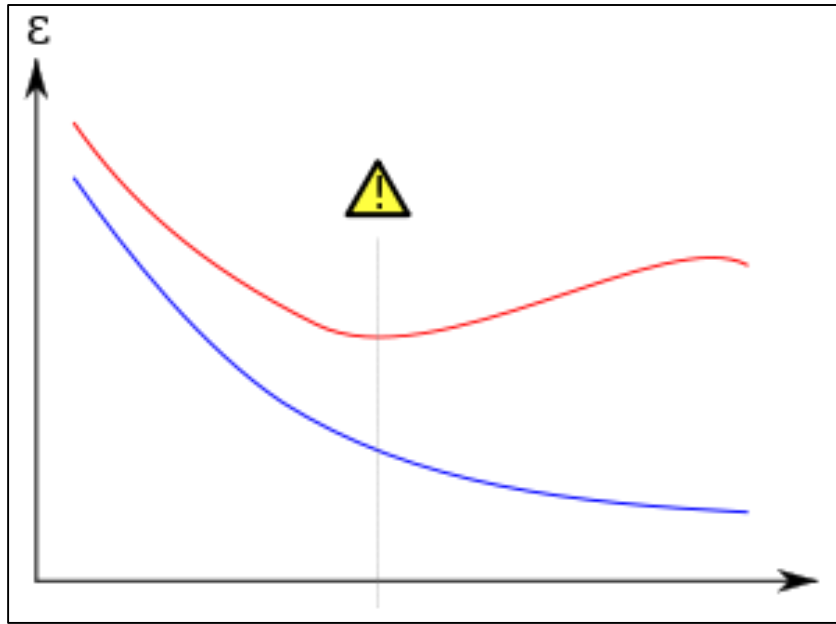
Data Scientist  
Jhoan Esteban Ruiz Borja Msc

### Sobreajuste (Overfitting)

En **aprendizaje automático** (**Machine Learning**), el **sobreajuste** (también es frecuente emplear el término en inglés **overfitting**) es el efecto de sobreentrenar un algoritmo de aprendizaje con unos ciertos datos para los que se conoce el resultado deseado. El algoritmo de aprendizaje debe alcanzar un estado en el que será capaz de predecir el resultado en otros casos a partir de lo aprendido con los datos de entrenamiento, generalizando para poder resolver situaciones distintas a las conocidas durante el entrenamiento. Sin embargo, cuando un sistema se entrena demasiado (se sobreentrena) o se entrena con datos extraños, el algoritmo de aprendizaje puede quedar ajustado a unas características muy específicas de los datos de entrenamiento que no tienen relación causal con la función objetivo. Durante la fase de sobreajuste el éxito al responder las muestras de entrenamiento sigue incrementándose mientras que su actuación con muestras nuevas va empeorando.



**Figura 1:** El emplear la línea verde como clasificador se adapta mejor a los datos con los que hemos entrenado al clasificador, pero está demasiado adaptado a ellos, de forma que ante nuevos datos probablemente arrojará más errores que la clasificación usando la línea negra.

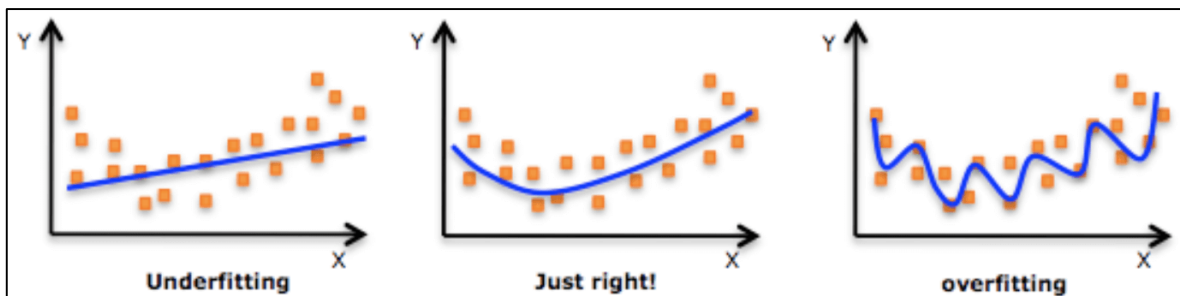


**Figura 2:** Sobreajuste/Sobreentrenamiento en aprendizaje supervisado (p. ej. una red neuronal). El error de entrenamiento se muestra en azul, mientras que el error de validación se muestra en rojo. Si el error de validación se incrementa mientras que el de entrenamiento decrece puede que se esté produciendo una situación de sobreajuste.

## Subajuste (Underfitting)

**Underfitting** ocurre cuando un modelo estadístico no puede capturar adecuadamente la estructura subyacente de los datos. Un modelo **underfitting** es un modelo donde faltan algunos parámetros o términos que aparecerían en un modelo correctamente especificado. El **underfitting** se produciría, por ejemplo, al ajustar un modelo lineal a datos no lineales. Tal modelo tenderá a tener un desempeño predictivo pobre.

**Underfitting**, la contrapartida del sobreajuste (**Overfitting**), ocurre cuando un modelo de aprendizaje automático no es lo suficientemente complejo como para capturar con precisión las relaciones entre las características de un conjunto de datos y una variable objetivo. Un modelo con underfitting da como resultados problemáticos o erróneos en los nuevos datos, o datos en los que no se entrenó, y muchas veces tiene un desempeño deficiente incluso en los datos de entrenamiento.



## **¿Por qué es importante el Underfitting?**

El uso de los modelos con underfitting para la toma de decisiones podría ser costoso para las empresas. Por ejemplo, un modelo con underfitting equipado puede sugerir que siempre puede realizar mejores ventas gastando más en marketing cuando, de hecho, el modelo no logra captar un efecto de saturación (en algún momento, las ventas se aplanarán sin importar cuánto más gaste en marketing). Si su empresa depende de ese modelo para determinar su presupuesto de marketing, gastará de más en marketing.