



II Matemáticas y Datos...

Dilema del sesgo y la varianza in a nutshell



El dilema del sesgo y la varianza es la dificultad más grande que existe en cualquier decisión (modelo) basada en la experiencia (datos): cuando un piloto de carreras está conociendo una pista complicada, debe decidir entre utilizar un vehículo difícil de controlar (varianza) y otro que no le permita maniobrar lo suficiente (sesgo).

Matemáticamente el dilema entre el sesgo y la varianza es la siguiente descomposición del error en un problema supervisado:

$$Error(\hat{Y}_{Datos}, Y) = Error(\hat{Y}_{Datos}, Y_{Prior}) + Error(Y_{Prior}, Y)$$

La primera versión de esta descomposición fue introducida para estudiar a las redes neuronales profundas respecto al error cuadrático medio, en este caso podemos interpretar al primer sumando como la varianza de los pesos durante el entrenamiento y al segundo como el ruido el cual tradicionalmente es gaussiano.

El prior en este caso será la arquitectura de la red neuronal.