

Dimensión V_C = Número de parámetros

Cantidad de datos necesarios = $10 d_{vc}$

$$E_{in}(h) \approx E_{out}(h) \Rightarrow M \gg 10 d_{vc}(H)$$

Podemos usar un modelo con menos parámetros para asegurar $E_{in}(h) \approx E_{out}(h)$, pero llegamos a sacrificar $E_{in}(h) \approx 0$ y si eso pasa solo podemos asegurar que no está aprendiendo

Para una cantidad masiva de datos, las redes neuronales mejoran mucho, pero no sirven con una pequeña cantidad de datos y es aquí donde los modelos clásicos

Investigar
TEORÍA
DE LA
INFORMACIÓN