

- Def.: crear y usar modelos 'aprendidos' de los datos

- Ej.:
- Predecir si un mail es spam
 - " si una transacción con tarjeta de crédito es fraudulenta.
 - " qué artículos le interesan más a un consumidor

- Modelos:
- supervisados: un ^{sub}conjunto de los datos está clasificado. Ej.: un conjunto de mails están clasificados de antemano como spam o no-spam
 - no-supervisados: no hay subconjuntos de datos clasificados de antemano.

- 2 riesgos:
- overfitting: modelos q' funcionan bien con los datos de entrenamiento pero q' se generalizan mal a nuevos datos. Esto se debe a q' el modelo 'aprende' ruido o q' 'aprende' a distinguir casos específicos de input (esto puede pasar con un modelo sobre-entrenado o muy rígido).

Underfitting: modelos q' ni siquiera se comportan bien con los datos de entrenamiento

- Entrenamiento: - En los modelos supervisados se puede utilizar el subconjunto de datos clasificados para entrenar el modelo.