

Aprendizaje automatizado

SELECCIÓN DE MODELOS

Gibran Fuentes-Pineda

Marzo 2023

- Demasiadas características pueden degradar el rendimiento de modelos
 - Maldición de la dimensionalidad
 - Atributos redundantes
 - Atributos irrelevantes

Extracción de características vs selección de atributos

- **Extracción de características:** mapea los atributos a un espacio de dimensiones menores
- **Selección de atributos:** elige un subconjunto de los atributos existentes
 - *Filtros:* evalúan el contenido de los atributos (por ej. distancia entre clases)
 - *Envolventes:* usan el clasificador para evaluar subconjuntos de atributos
 - *Híbridos:* tratan de combinar las ventajas de los filtros y los envolventes

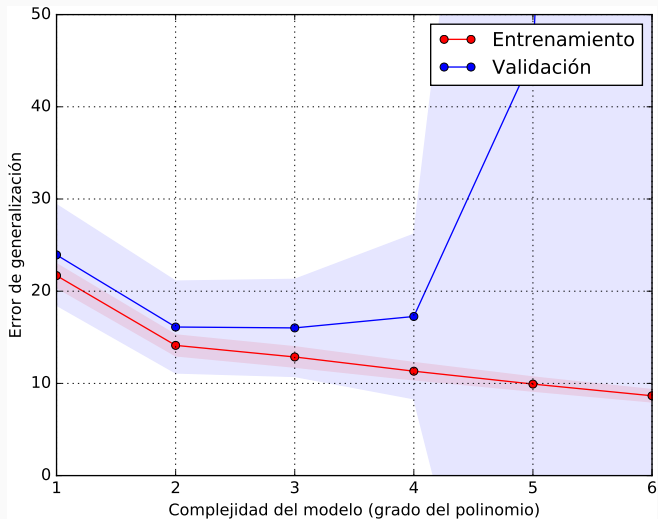
¿Por qué reducir el número de atributos?

- Menos efectos de la maldición de la dimensionalidad
- Menos espacio y mediciones
- Más rápido de entrenar y ejecutar
- Más fácil de interpretar y visualizar

¿Cómo elegimos los atributos más adecuados?

- Búsqueda óptima de subconjuntos es intratable
- **Selección hacia adelante:** se va añadiendo incrementalmente el atributo que disminuya más el error
- **Selección hacia atrás:** se va eliminando decrementalmente el atributo que aumente más el error

Sesgo vs varianza



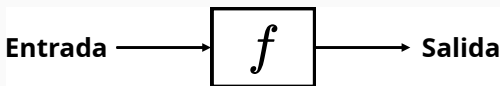
Criterio de información bayesiana (BIC)

- Es posible incrementar la verosimilitud de cualquier modelo haciéndolo más complejo a costo de posible sobre-ajuste
- BIC es un criterio que penaliza modelos con muchos parámetros

$$BIC = -2 \cdot \log(\text{máx likelihood}) + \log(n) \cdot d$$

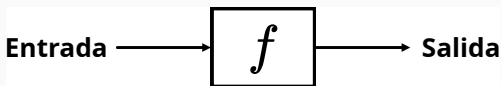
- Existen distintas estrategias para elegir valores apropiados de hiperparámetros que están basadas en evaluar el desempeño de los modelos usando validación cruzada
- Ejemplos
 - Búsqueda de rejilla
 - Búsqueda aleatoria
 - Algoritmos evolutivos
 - Optimización bayesiana

- ¿Es posible aprender cualquier tarea (función f)? ¿Es necesario el conocimiento a priori?



¹D. Wolpert. The Lack of A Priori Distinctions between Learning Algorithms, *Neural Computation*, pp. 1341–1390.

- ¿Es posible aprender cualquier tarea (función f)? ¿Es necesario el conocimiento a priori?



- *No existe la comida gratis*¹
 - Sólo es posible aprender de forma eficiente un pequeño subconjunto de todas las tareas posibles
 - El *sesgo inductivo* ayuda a aprender ciertas tareas

¹D. Wolpert. The Lack of A Priori Distinctions between Learning Algorithms, *Neural Computation*, pp. 1341–1390.