

#8 Selección de características (Feature Engineering)

¿En qué consiste la selección de características de un modelo de machine learning?

```
X = df[["variable_1", "variable_2"]]
```

```
y = df["variable_objetivo"]
```

Respuesta/=

La selección de características en un modelo de machine learning implica identificar y elegir las variables o características más relevantes para predecir la variable objetivo. En el código que proporcionaste, parece que estás seleccionando dos características, "variable_1" y "variable_2", del DataFrame df para usarlas como entrada (X) en tu modelo de machine learning, mientras que "variable_objetivo" se utiliza como la variable objetivo (y).

La selección de características es importante por varias razones:

Reducción de dimensionalidad: Al seleccionar las características más relevantes, puedes reducir la dimensionalidad de los datos, lo que puede mejorar la eficiencia computacional y reducir el riesgo de sobreajuste.

Mejora del rendimiento del modelo: Al eliminar características irrelevantes o redundantes, puedes mejorar la precisión y la generalización del modelo.

Interpretación del modelo: Al seleccionar características significativas, el modelo resultante es más fácil de interpretar y entender.

Existen diversas técnicas para realizar la selección de características, que van desde métodos estadísticos simples hasta algoritmos más avanzados. Algunos ejemplos de técnicas comunes incluyen:

Prueba estadística univariante: Evalúa la relación entre cada característica y la variable objetivo utilizando pruebas estadísticas como la prueba t o la prueba F.

Selección de características basada en modelos: Entrena un modelo y utiliza su rendimiento para evaluar la importancia de cada característica. Algunos ejemplos son la importancia de características en árboles de decisión o coeficientes en modelos lineales.

Selección de características recursivas: Entrena un modelo y elimina iterativamente las características menos importantes, repitiendo el proceso hasta que se alcanza un criterio de parada.

Selección de características basada en la información mutua: Calcula la cantidad de información que una característica proporciona sobre la variable objetivo.

La elección de la técnica de selección de características adecuada depende del problema específico y de las características de los datos. Es importante experimentar con diferentes

métodos y evaluar su rendimiento utilizando técnicas de validación cruzada u otras métricas de evaluación del modelo.

#####

Eva Bernal: Pegar tu enlace aqui:

#####

#####

Manuel Murillo: Pegar tu enlace aquí:

https://github.com/manuelmurillo1984/p3_pull_requests/blob/ac15dea77fc9a7221c17b6f5418a8e10ed217110/manuelmurillo.pdf

#####

#####

Humberto Sánchez: Pegar tu enlace aquí:

#####