

Support vector machines

Fabián Villena

Enero 2024

Se le pide seleccionar el mejor modelo a la hora de predecir el tipo de medicamento adecuado para un paciente.

En particular, trabajaremos con un conjunto de datos que contiene información sobre la clasificación de medicamentos basada en la información general del paciente y su diagnóstico. La idea es construir modelos de aprendizaje automático para predecir el tipo de fármaco (*Drug*) que podría ser adecuado para el paciente.

Los resultados esperados son que puedan realizar un flujo clásico al momento de crear modelos de aprendizaje supervisado.

<https://github.com/fvillena/biocompu/blob/2023/data/drug.csv>

Preguntas

Responda las siguientes preguntas en un *Jupyter Notebook* con código desarrollado en el lenguaje de programación Python.

1. Utilizar la técnica de *one-hot encoding* para transformar variables categóricas a numéricas.
2. Separe el conjunto de datos en un subconjunto de entrenamiento y prueba
3. Entrene múltiples modelos de clasificación basados en Support Vector Machines con distintos hiperparámetros.
4. Prediga sobre el subconjunto de prueba utilizando cada uno de los modelos ajustados.
5. Calcule al menos dos métricas de rendimiento para cada uno de los modelos de clasificación.
6. Seleccione la mejor combinación de hiperparámetros.