

Ejercicios para practicar: KNN

El conjunto de datos VentaViviendas contiene información sobre el precio de venta de una serie de viviendas, junto con las características básicas de las mismas. Las variables contenidas en el fichero son:

Variable	Descripción
price (var. dependiente)	Precio de venta de la vivienda
bathrooms	Número de baños (los medios se refieren a aseos)
superf	Superficie del interior de la vivienda (en m ²)
garden	Superficie del jardín (en hectáreas)
floors	Número de plantas
view	¿Tiene buenas vistas? (1: sí, 0: no)
condition	Estado de la vivienda (de 1 a 3, siendo 1 el mejor estado)
antig	Antigüedad (en años) de la vivienda
renovated	¿La vivienda ha sido reformada? (1: sí, 0: no)
lat, long	Coordenadas de latitud y longitud de la vivienda

Los ejercicios consisten en lo siguiente (recuerda que debes llevarlos a cabo sobre un archivo rmd para posteriormente generar el *html* y subirlo al campus virtual):

1. Carga los datos en el entorno de Rstudio a través de la función *readRDS*. Utilizando el código que consideres (y los datos disponibles), indica qué precio estimarías que tiene una vivienda.
2. Realiza una partición del conjunto de datos en entrenamiento (80%) y prueba (20%).
3. Genera un primer modelo KNN para la variable price con un k igual a 5 y la distancia de Manhattan. Imprime las matrices D y CL, explica qué contienen y relaciónalo con el funcionamiento del modelo KNN. Así mismo, calcula el R² prueba.
4. Utilizando validación cruzada, haz un análisis preliminar del k y tipo de distancia óptimo. Comenta los resultados y determina la "mejor" combinación, justificando porqué consideras que es la mejor.
5. Utilizando esa combinación óptima, lleva a cabo una selección de variables. De nuevo, comenta los resultados y concluye cuál es el conjunto óptimo de variables input.
6. Para la combinación óptima de variables, busca el valor óptimo de k (cerrando el conjunto de valores teniendo en cuenta los resultados del apartado 4) y del tipo de distancia. Comenta los resultados e indica el resultado.
7. Construye el mejor modelo y calcula el R² prueba. Usando la información de la validación cruzada anterior, comenta si se trata de un modelo estable.
8. Obtén el PDP para dos variables (una cuantitativa y una cualitativa), juntas y por separado. Comenta los resultados.

Nota: como algunos procesos tardan en ejecutarse, te recomiendo que comentes los resultados de cada ejercicio mientras se ejecuta el siguiente.