

Preguntas.

- a) Capturing patterns from data is called fitting or training the model.
- b) The data used to fit the model is called the training data.
- c) After the model has been fit, you can apply it to new data to predict prices of additional homes.
- d) You predict the price of any house by tracing through the decision tree, always picking the path corresponding to that house's characteristics.
- e) The point at the bottom where we make a prediction is called a leaf.
- f) Explica lo que representa cada fila de la siguiente tabla.

Los resultados muestran 8 números para cada columna en su conjunto de datos original. El primer número, el recuento, muestra cuántas filas tienen valores no faltantes.

Los valores perdidos surgen por muchas razones. Por ejemplo, el tamaño de la segunda habitación no se recogería al inspeccionar una casa de 1 habitación. Volveremos al tema de los datos faltantes.

El segundo valor es la media, que es el promedio. Debajo de eso, std es la desviación estándar, que mide la extensión numérica de los valores.

Para interpretar los valores mínimo, 25%, 50%, 75% y máximo, imagine ordenar cada columna del valor más bajo al más alto. El primer valor (el más pequeño) es el mínimo. Si recorre un cuarto de camino en la lista, encontrará un número que es mayor que el 25% de los valores y menor que el 75% de los valores. Ese es el valor del 25% (pronunciado "percentil 25"). Los percentiles 50 y 75 se definen de forma análoga, y el máximo es el número más grande.

- g) The column we want to predict, which is called the prediction target
- h) The columns that are inputted into our model (and later used to make predictions) are called features
- i) The steps to building and using a model are:

Define: ¿Qué tipo de modelo será? ¿Un árbol de decisiones? ¿Algún otro tipo de modelo? También se especifican algunos otros parámetros del tipo de modelo.

Fit: Capture patrones de los datos proporcionados. Este es el corazón del modelaje.

Predict: Justo como suena

Evaluate: Determine qué tan precisas son las predicciones del modelo.

j) Observa cómo es la predicción de los precios tomando en cuenta cinco registros. NOTA: En realidad no son las primeras cinco en orden de aparición puesto que se excluyeron los registros que tienen celdas vacías en el archivomelb_data.csv ¿cuál sería el MAE para los datos predichos?

[1035000. 1465000. 1600000. 1876000. 1636000.]

k) ¿Cómo se define el Error Medio Absoluto (MAE Mean Absolute Error)

Es una métrica que mide la calidad del modelo, donde se indica que el error es igual al valor actual menos el valor de la predicción.

l) ¿Cuál es el MAE de tu entrenamiento con cinco registros?

666.6666666666666

m) Después de dividir el conjunto de datos en las variables de entrenamiento y validación (train_X, val_X, train_y, val_y), así como después de entrenar nuevamente, cuál es el MAE que obtienes?

304900.0