

Resumen del capítulo: Pasar a la regresión

Error cuadrático medio

La métrica de evaluación más comúnmente usada para las tareas de regresión es el **error cuadrático medio o ECM**.

Para encontrar el *ECM*, primero debes calcular el error de cada observación:

$$\text{Error de observación} = \text{Predicción del modelo} - \text{Respuesta correcta}$$

Calcula el *ECM* usando esta fórmula.

$$\text{ECM} = \frac{\text{Suma de los cuadrados de los errores de observación}}{\text{Número de observaciones}}$$

Vamos a analizar estos cálculos:

1. El error de observación muestra el grado de discrepancia entre la respuesta correcta y la predicción. Si el error es mucho más grande que cero, el modelo ha sobrevalorado el apartamento; si es mucho menor que cero, entonces el modelo lo ha infravalorado.

2. No tendría sentido sumar los errores tal y como están, ya que los positivos anularían los negativos. Para hacer que todos cuenten, necesitamos deshacernos de los signos elevando al cuadrado cada uno de ellos.

3. Encontramos el promedio para obtener datos para todas las observaciones.

El *ECM* debe ser lo más bajo posible.

Cálculo de ECM

Para calcular el error medio cuadrático, importa la función `mean_squared_error()` del módulo `sklearn.metrics`.

```
from sklearn.metrics import mean_squared_error

mse = mean_squared_error(answers, predictions)
```

Como resultado, obtendrás unidades cuadradas (por ejemplo, "dólares cuadrados"). Para conseguir una métrica de evaluación en las unidades regulares, encuentra la raíz cuadrada de *ECM*. Entonces, obtendrás *RMSE* (*raíz del error cuadrático medio*):

```
rmse = mse ** 0.5
```

Regresión del árbol de decisión

Para tareas de regresión, los árboles de decisión se entrenan de manera similar a la clasificación, pero no predicen una clase, sino un número.

El árbol de decisión para tareas de regresión se llama `DecisionTreeRegressor` y está en el módulo `sklearn.tree`.

```
from sklearn.tree import DecisionTreeRegressor

model = DecisionTreeRegressor(random_state=54321)
```

Regresión lineal

La regresión lineal es similar a la regresión logística en varios aspectos. Su nombre proviene del álgebra lineal. La regresión lineal es menos propensa al sobreajuste porque no tiene muchos parámetros.

```
from sklearn.linear_model import LinearRegression  
  
model = LinearRegression()
```