Hasan Akhtar Machine Learning (Linear Regression) example

An approach to a simple machine learning problem about predicting the orders required for 4 different drinks based on past data and any current factors such as promotions, weather conditions, etc.

TASK:

Create a solution for the above mentioned problem using a linear regression model and fine tune it so that it can accurately predict future sales of these drinks for order stocks.

Training Data:

The training data is split as follows for each of the 4 drinks. Below are given 7 examples:

The actual training data is set out below as screenshots and the files are as follows:

PepsiSales.csv

FantaSales.csv

SpriteSales.csv

RubiconSales.csv

PepsiSales.csv

FantaSales.csv

SpriteSales.csv

RubiconSales.csv

The Code:

Lines 1-7 import the modules/libraries required to perform the machine learning

Lines 9-13 read from each file individually using the read\_csv method of the pandas module/library. These files contain the past data that has been stored.

Lines 15-17 spit the data into the Features (X) and Target variable (y)

Lines 19-21 use string methods and regular expressions to get the start time and end time from the date time string. These are required for the training data.

Line 24 uses a similar method as above to retrieve the day of the week from the date time string. This is also required for the training data.

Lines 26-28 convert the start time and end time to hour objects for the training data

Line 30 creates an updated version of the X array for the features of the dataset

Line 33 uses the train\_test\_split method to split the dataset into 80% for training and 20% for testing. This is essential to make sure the model can be evaluated after training

Lines 35-37 use machine learning methods to train the linear regression model

Line 40 evaluates the model using the test data

The same procedure as before is carried out for the second dataset

Then for the third dataset as well

Then finally the fourth and final dataset

The next few lines allow the user to enter information so that the model can predict how many drinks of a certain type will be sold. From line 134-145 an if-elif-else structure is used to determine which drink the model needs to evaluate. Then the dateTime is formatted and split into the start\_hour, end-hour and day\_of\_week. The other features are also inputted thereafter.

Finally the data entered by the user is used to make a somewhat accurate prediction and this is outputted to the user.

TESTS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date/Time | Temperature (°C) | Promotion (Y/N) | Customers | Drinks sold |
| Mon 9-10am | 21 | N | 152 | 241 |
| Mon 3-4pm | 30 | N | 362 | 578 |
| Tue 1-2pm | 15 | N | 466 | 479 |
| Tue 5-6pm | 18 | N | 422 | 453 |
| Thu 11-12pm | 16 | Y | 327 | 438 |
| Fri 2-3pm | 24 | N | 216 | 269 |
| Sat 4-5pm | 27 | Y | 265 | 380 |