Splitting Dependent And Independent Columns

UNIVERSITY ADMIT ELIGIBILITY PREDICTOR

TEAM ID : PNT2022TMID36734

We need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

To read the columns, we will use **iloc** of pandas (used to fix the indexes for selection) which takes two parameters — [row selection, column selection].

Let’s split our dataset into independent and dependent variables.

```
In [20]:  1 x=data.iloc[:,0:7].values
          2 x

Out[20]: array([[337.  , 118.  ,   4.  , ...,   4.5 ,   9.65,   1.  ],
                [324.  , 107.  ,   4.  , ...,   4.5 ,   8.87,   1.  ],
                [316.  , 104.  ,   3.  , ...,   3.5 ,   8.  ,   1.  ],
                ...,
                [330.  , 116.  ,   4.  , ...,   4.5 ,   9.45,   1.  ],
                [312.  , 103.  ,   3.  , ...,   4.  ,   8.78,   0.  ],
                [333.  , 117.  ,   4.  , ...,   4.  ,   9.66,   1.  ]])
```

```
In [22]:  1 y=data.iloc[:,7:].values
          2 y

                [0.76],
                [0.44],
                [0.46],
                [0.54],
                [0.65],
                [0.74],
                [0.91],
                [0.9 ],
                [0.94],
                [0.88],
                [0.64],
                [0.58],
                [0.52],
                [0.48],
                [0.46],
                [0.49],
                [0.53],
                [0.87],
                [0.91],
```

Let's Check the shape of X and Y.

Splitting Dependent And Independent Columns

```
In [23]:  1 x.shape

Out[23]: (400, 7)

In [24]:  1 y.shape

Out[24]: (400, 1)
```

You can see in x we have 400 rows with 7 columns and y has 1 column with the same number of rows.