Assignment No. 4

Title:

Execute feature scaling on given dataset

Theory:

Feature scaling is a technique to standardize the independent features present in data in a fixed range. It is performed during data pre-processing to in data in fixed range. If feature scaling is not done, then ML algorithm tends to weigh greater values, higher and consider smaller values as lower values, regardless of unit of values

Example:

If our algorithm is not using feature scaling method then it can consider value 3000 meters to be greater than 5 km but that's actually not true and in this case, algorithm will give wrong predictions. So we use feature scaling to bring all same magnitudes & thus, tackle this issue.

Techniques to perform feature scaling

Consider two most important ones:

Min-Max normalization: This technique re-scales feature or observation value with distribution value between 0 and 1

$$X_{new} = \frac{X_i - min(x)}{max(x) - min(x)}$$

standardization: It is very effective technique which re-scales a feature value so that distribution with 0 mean value & variance equals to 1

$$X_{new} = \frac{X_i - X_{mean}}{S.D}$$

	A	B	C	D
i)	country	Age	Salary	purchased
ii)	France	44	72000	0
iii)	Spain	27	48000	1
iv)	Germany	30	54000	0
v)	Spain	38	61000	0
vi)	Germany	40	1000	1
vii)	France	35	58000	1
viii)	Spain	78	52000	0
ix)	France	48	79000	1
x)	Germany	50	83000	0
xi)	France	37	67000	1
	~~Spain~~			

conclusion:

Thus, we have studied feature scaling on given dataset.