Date
Page

* all about StandardScaler() → feature scalling and when we have to perform it.
* and the difference between scalled and non-scalled data.

- feature scaling, such standardization (z-score scaling) or min-max scaling, is intended for continuous numerical features.

- Scaler.fit_transform and scaler.transform are two methods used with scalling objects from scikit-learn library

Scalar.fit_transform :	scaler.transform
- used to both fit the scaling parameters to your training data and transform it in single step	- used to transform a dataset using scaling parameters that were previously fitted using ('fit' or fit_transform) on another dataset.
- It computes the mean and standard deviation of your training data and then scales the data based on thes computed parameters	- it does not recompute the parameters. it directly applies the scalling based on the parameters already determined during the fitting process.
- It is typically used on the training data.	- It is commonly used on the test or validation data to ensure that the scaling is applied as on training data, maintaining consistancy.

* on test data set use only transform().
- it use the ~~parameter~~ scaling parameter from the Training set to scale the test set.
- This is crucial to avoid data leakage and to maintain the sepration between the training and test sets.

```python
from sklearn.preprocessing import StandardScaler

# Create a toy dataset
data = [[1, 2], [2, 3], [3, 4]]

# Create a StandardScaler
scaler = StandardScaler()

# Fit and transform the data in one step (common for training data)
scaled_data_train = scaler.fit_transform(data)

# New data (e.g., test data)
new_data = [[4, 5], [5, 6]]

# Use the previously fitted scaler to transform the new data (common for test data)
scaled_new_data = scaler.transform(new_data)

print("Scaled Training Data:")
print(scaled_data_train)

print("Scaled New Data:")
print(scaled_new_data)
```

Output :-

```
Scaled Training Data:
[[-1.22474487 -1.22474487]
 [ 0.          0.        ]
 [ 1.22474487  1.22474487]]
Scaled New Data:
[[2.44948974 2.44948974]
 [3.67423461 3.67423461]]
```

- with above example we get to know how we can ~~get~~ Compute the scaling

∴ scaled training data

date		
	1	2
	2	3
	3	4

-1.23	

$\mu = \frac{1+2+3}{3}$

$\boxed{\mu = 2}$

$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$

$= \sqrt{\frac{1+0+1}{3}}$

$= \sqrt{0.66}$

$\boxed{\sigma = 0.8124}$

like that Calculate for all

for Scaling

Z-score formula

Standardized value $(Z) = \frac{(X - \mu)}{\sigma}$

$= \frac{1-2}{0.8124} = -1.230$

Now. New data

4	5
5	6

∴ Scaled New data

2.461	.

Calculate for all.

Compute scaling with $\mu = 2$, $\sigma = 0.8124$

∴ for 4 → $\frac{4-2}{0.8124} = \frac{2}{0.8124} = 2.461$