Assignment 4

Name: Mohammed Varaliya

Roll No: 54

Question

1. **Why don't we apply standardization on** `y_train` **?**

Apply Standardization

$$z = \frac{x - \mu}{\sigma}$$

```
[28]  sc = StandardScaler()
```

```
[29]  X_train_scaled = sc.fit_transform(X_train)
      X_test_scaled = sc.transform(X_test)
```

1. **Why don't we apply standardization on** `y_train` **?**

    a. We don't apply standardization on `y_train` in most cases because:

        i. **Explanation:**

            1. Standardization is primarily used for features (X) to ensure they have a mean of 0 and a standard deviation of 1, making them comparable and improving the performance of machine learning models, especially those sensitive to the scale of input features (e.g., linear regression, SVMs, neural networks).

        ii. **Reasoning:**

            1. The target variable (`y_train`) typically represents the output we are trying to predict. Standardizing `y_train` could distort the actual values that the model needs to predict. For regression problems, we usually keep `y_train` in its original scale so that the model's predictions are also in the same scale as the target, allowing for easier interpretation and proper evaluation. Only in cases like certain types of regression models (e.g., when working with multiple output variables or complex transformations), would you consider scaling `y_train` , but it's generally avoided.