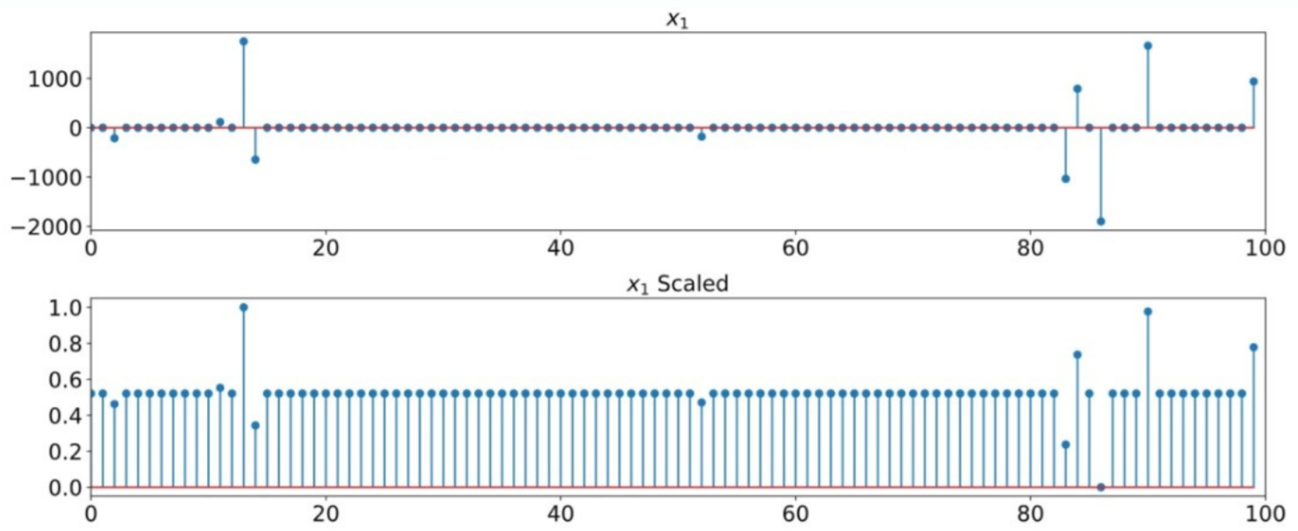

The figure shows a values of a feature (x_1) before and after scaling. What is the scaling method used? [5 marks]

- [ ] a. Standard scaling
- [x] b. Min-max scaling
- [ ] c. Max-abs scaling

output range [0-1] → min-max

Data preprocessing example

- https://scikit-learn.org/stable/modules/preprocessing.html

1. Standardization: scale the features of a dataset to have zero mean and unit variance.
2. Scaling features to a range e.g., between 0 and 1
   - Min max scalar → [0, 1]
   - Max Abs Scaler → [-1, 1]

$$\text{Standardization}(x) = \frac{x - \text{mean}(x)}{\text{std}(x)}$$

$$\text{MinMaxScaler}(x) = \frac{x - \min(x)}{\max(x) - \min(x)}$$

$$\text{MaxAbsScaler}(x) = \frac{x}{\max(|x|)}$$

If outliers are there, will it work?

Suggestions?

Given a data set of observations $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)^T$ and two machine learning models namely model 1 and 2. The likelihood values for model 1 are expressed as $p(\boldsymbol{x}_1|\boldsymbol{\theta}_1) = 0.5$, $p(\boldsymbol{x}_2|\boldsymbol{\theta}_1) = 0.7$ and $p(\boldsymbol{x}_3|\boldsymbol{\theta}_1) = 0.4$, while for model 2 they are expressed as $p(\boldsymbol{x}_1|\boldsymbol{\theta}_2) = 0.6$, $p(\boldsymbol{x}_2|\boldsymbol{\theta}_2) = 0.5$ and $p(\boldsymbol{x}_3|\boldsymbol{\theta}_2) = 0.4$. Here, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ represents the parameters of the model 1 and 2, respectively. Which model would be chosen based on the utilization of maximum likelihood analysis? [15 marks]

☐ a. Model 2

☑ b. Model 1

Model 1 Likelihoods:

- $p(x1|\theta_{01}) = 0.5$
- $p(x2|\theta_{01}) = 0.7$
- $p(x3|\theta_{01}) = 0.4$

These values represent the likelihood of observing each data point x_i under model 1, with parameters θ_{01}.

Model 2 Likelihoods:

- $p(x1|\theta_{02}) = 0.6$
- $p(x2|\theta_{02}) = 0.5$
- $p(x3|\theta_{02}) = 0.4$

These values represent the likelihood of observing each data point x_i under model 2, with parameters θ_{02}.

Maximum Likelihood Decision:

For each model, the likelihood of the entire dataset $X = (x1, x2, x3)$ is calculated as the product of individual likelihoods: $L(\theta) = p(x1|\theta) \cdot p(x2|\theta) \cdot p(x3|\theta)$

Model 1:

$$L(\theta_{01}) = 0.5 \times 0.7 \times 0.4 = 0.14$$

Model 2:

$$L(\theta_{02}) = 0.6 \times 0.5 \times 0.4 = 0.12$$

model 1 likelyhood > model 2 likelyhood

Consider the linear regression model $y = w_0 + w_1 x_1 + w_2 x_2 = \boldsymbol{w}^T \boldsymbol{x}$. Sum of squared errors (SSE) and total sum of squares (TSS) of this model for two datasets are given in the Table. For which dataset this model performs better? [15 marks]

Table 5: SSE and TSS of linear regression model.

	Dataset A	Dataset B
SSE= $\sum_{i=1}^{N}(y_i - \boldsymbol{w}^T \boldsymbol{x}_i)^2$	1	8
TSS= $\sum_{i=1}^{N}(y_i - \bar{y}_i)^2$	5	80
Number of data samples (N)	10000	10000

$$R^2 = 1 - \frac{SSE}{TSS}$$

Given Data:

- **Dataset A:**
  - SSE = 1
  - TSS = 5
- **Dataset B:**
  - SSE = 8
  - TSS = 80

Calculating R^2:

1. **For Dataset A:**

$$R_A^2 = 1 - \frac{SSE_A}{TSS_A} = 1 - \frac{1}{5} = 1 - 0.2 = 0.8$$

2. **For Dataset B:**

$$R_B^2 = 1 - \frac{SSE_B}{TSS_B} = 1 - \frac{8}{80} = 1 - 0.1 = 0.9$$

$R_A^2 < R_B^2$ model B performs better.

1. **SSE (Sum of Squared Errors)**:

observed value

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

predicted value

where y_i are the observed values, $\hat{y}_i$ are the predicted values, and n is the number of observations.

2. **TSS (Total Sum of Squares)**:

observed value

$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

mean

where $\bar{y}$ is the mean of the observed values.

3. **R-squared**:

$$R^2 = 1 - \frac{\text{SSE}}{\text{TSS}}$$

high R^2 value says a better model