BCSE209P- Machine Learning Lab

Digital Assignment 2

Problem Statement:

Suppose you have height and weight data for a group of people. For example: Heights are in feet, like 6.5, and weight is in grams, like 80000. In many machine learning situations, you want to normalize the data — scale the data so that the values in different columns have roughly the same magnitude so that large values (like the weight) don't overwhelm smaller values (like the heights). Create a raw data of minimum 40 records of height and weight in above mentioned format and use Min-Max Normalization to normalize the weights in the range from as well as use Z-score to normalize the weights.

Concept to be applied:

Feature scaling becomes necessary when dealing with datasets containing features that have different ranges, units of measurement, or orders of magnitude. In such cases, the variation in feature values can lead to biased model performance or difficulties during the learning process.

There are several common techniques for feature scaling, including standardization, normalization, and min-max scaling. These methods adjust the feature values while preserving their relative relationships and distributions.

Min-Max Scaling is a simple method of normalization that scales the values between 0 and 1. It works by subtracting the minimum value from each value in the dataset and then dividing by the range of the dataset (i.e., maximum value minus minimum value).

X_norm = (X - X.min()) / (X.max() - X.min())

where `X` is the original dataset, `X_min` is the minimum value of `X`, and `X_max` is the maximum value of `X`.

Z-score is a variation of scaling that represents the number of standard deviations away from the mean. You would use z-score to ensure your feature distributions have mean = 0 and std
= 1. It's useful when there are a few outliers, but not so extreme that you need clipping.

X_norm = (X - X.mean()) / X.std()

where `X` is the original dataset, `X_mean` is the mean value of `X`, and `X_std` is the standard deviation of `X`.