

Mean Centring and Feature Scaling

Tuesday 21 May 2024 1:03 AM

■ Mean-Centering and Feature-Scaling (Normalization or Standardization)

- Mean centering is a technique that shifts data closer to the origin. Since the regression line for the shifted data passes through the origin, we only need to estimate w without estimating b . Then we calculate b using w and the average value of the original data.
- Normalization or Standardization is a technique that not only performs mean centering but also feature scaling. In multivariate regression, if the scales of x_1 and x_2 are significantly different, this can also affect the estimate of w and b , so we need to standardize the scales of all x .
- Normalization is necessary, especially when using regularization. Without normalization, the regularization term may unfairly impose a greater penalty on some coefficients than on others.
- Estimate w' using the normalized data and convert it to the slope of the original data using the formula at the bottom right. Then calculate b using w and the average value of the original data. This is equivalent to mean centering.
- Mean centering or normalization is applied only to the training data. It does not need to be applied to the validation or test data.

