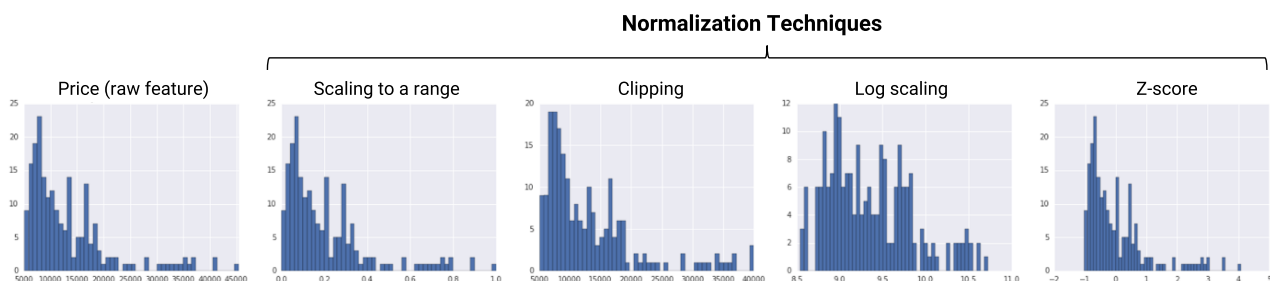# Normalization

The goal of normalization is to transform features to be on a similar scale. This improves the performance and training stability of the model.

Normalization Techniques at a Glance

Four common normalization techniques may be useful:

- scaling to a range

- clipping

- log scaling

- z-score

The following charts show the effect of each normalization technique on the distribution of the raw feature (price) on the left. The charts are based on the data set from 1985 Ward's Automotive Yearbook that is part of the UCI Machine Learning Repository under Automobile Data Set (https://archive.ics.uci.edu/ml/datasets/automobile).

Figure 1. Summary of normalization techniques.

Scaling to a range

Recall from MLCC (/machine-learning/crash-course/representation/cleaning-data) that **scaling** (/machine-learning/glossary#scaling) means converting floating-point feature values from their natural range (for example, 100 to 900) into a standard range—usually 0 and 1 (or sometimes -1 to +1). Use the following simple formula to scale to a range:

$$x' = (x - x_{min})/(x_{max} - x_{min})$$

Scaling to a range is a good choice when both of the following conditions are met:

- You know the approximate upper and lower bounds on your data with few or no outliers.

- Your data is approximately uniformly distributed across that range.

A good example is age. Most age values falls between 0 and 90, and every part of the range has a substantial number of people.

In contrast, you would *not* use scaling on income, because only a few people have very high incomes. The upper bound of the linear scale for income would be very high, and most people would be squeezed into a small part of the scale.

Feature Clipping

If your data set contains extreme outliers, you might try feature clipping, which caps all feature values above (or below) a certain value to fixed value. For example, you could clip all temperature values above 40 to be exactly 40.

You may apply feature clipping before or after other normalizations.

Formula: Set min/max values to avoid outliers.
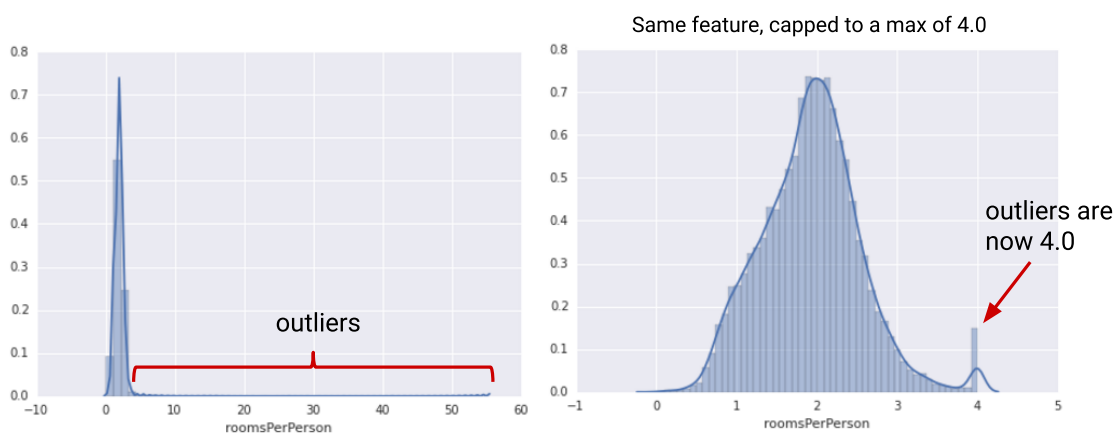
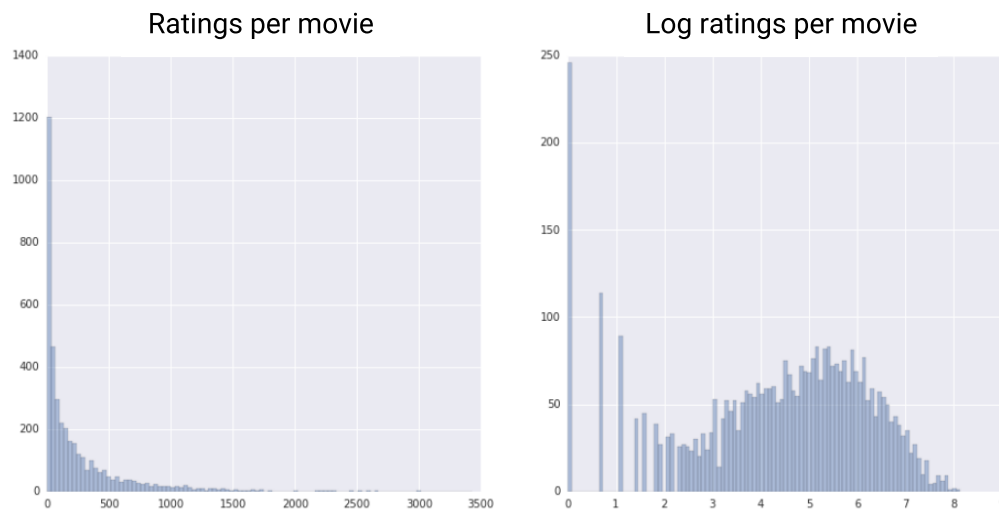**Figure 2. Comparing a raw distribution its clipped version.**

Another simple clipping strategy is to clip by z-score to +-Nσ (for example, limit to +-3σ). Note that σ is the standard deviation.

Log Scaling

Log scaling computes the log of your values to compress a wide range to a narrow range.

$$x' = log(x)$$

Log scaling is helpful when a handful of your values have many points, while most other values have few points. This data distribution is known as the *power law* distribution. Movie ratings are a good example. In the chart below, most movies have very few ratings (the data in the tail), while a few have lots of ratings (the data in the head). Log scaling changes the distribution, helping to improve linear model performance.

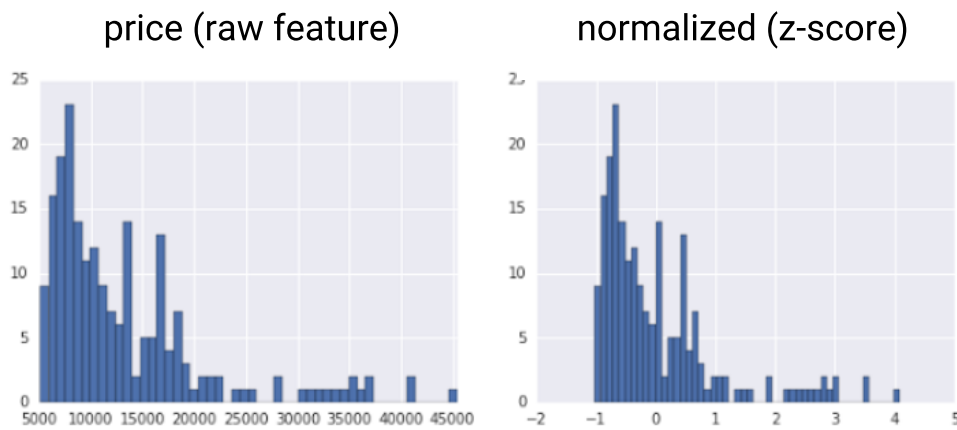**Figure 3. Comparing a raw distribution to its log.**

Z-Score

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.

The formula for calculating the z-score of a point, *x*, is as follows:

$$x' = (x - \mu)/\sigma$$

μ is the mean and σ is the standard deviation.

price (raw feature)　　　normalized (z-score)

Figure 4. Comparing a raw distribution to its z-score distribution.

Notice that z-score squeezes raw values that have a range of ~40000 down into a range from roughly -1 to +4.

Suppose you're not sure whether the outliers truly are extreme. In this case, start with z-score unless you have feature values that you don't want the model to learn; for example, the values are the result of measurement error or a quirk.

Summary

est normalization technique is one that empirically works well, so try new ideas if you think they'll work we
eature distribution.

Normalization Technique	Formula	When to Use
Linear Scaling	$x' = (x - x_{min})/(x_{max} - x_{min})$	When the feature is more-or-less uniformly distributed across a fixed range.
Clipping	if x > max, then x' = max. if x < min, then x' = min	When the feature contains some extreme outliers.
Log Scaling	x' = log(x)	When the feature conforms to the power law.
Z-score	x' = (x - μ) / σ	When the feature distribution does not contain extreme outliers.

:rms:

aling (/machine-learning/glossary#scaling)   •   normalization (/machine-learning/glossary#normalizat

Previous

← **Transforming Numeric Data** (/machine-learning/data-prep/transform/transform-numeric)

Next

Bucketing (/machine-learning/data-prep/transform/bucketing)   →