Usually the inputs to neural networks and other models are classed into quantitative and categorical values. The theoretical range of one quantitative input could be unbounded (-∞,+∞), another might be restricted to values between 0 and 1 or -1 and 1. Categorical variables are encoded to resemble quantitative variables: For example, a variable might be binned: 0-9 becomes 9, 10-19 becomes 10, etc. In all these cases, the inputs are made to lie on a straight number line.

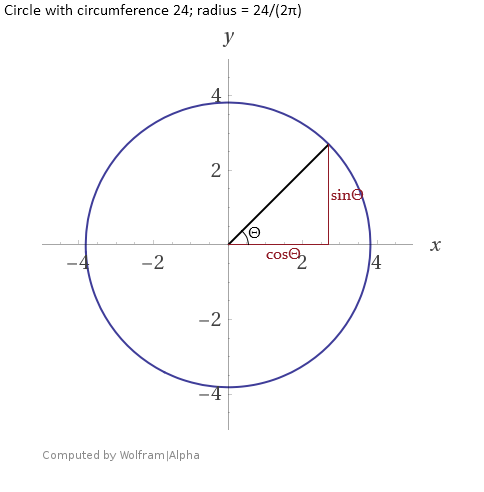
Clustering and all algorithms that use a distance metric (KNN, DBSCAN, K-Means) need a way to measure the distance between points of data. The most common distance metric is the Euclidean distance, which gives the length of the shortest straight path between a pair of points in space, without any turns. Periodic variables have

This article assumes some knowledge on training data preparation; things like scaling, encoding, imputation, and feature engineering.

Many cases involve *periodic* (or *cyclical*) features. These are features that logically have a loop in their values; for example, time of day (24-hour cycle), seasons, angles of rotating bodies, virtual wraparound spaces, and most anything sinusoidal. Acceptable methods for handling these values are dependent on the algorithms employed.

The most common solution to use periodic data in a neural network is to transform each periodic attribute to a two-dimensional circle, so that the inputs are the sine and cosine of the angle.

|  |  |  |
| --- | --- | --- |
| Time | Input A, sinΘ\* | Input B, cosΘ\* |
| Midnight | sin(0) = 0 | cos(0) = 1 |
| 3 AM | sin(π/4) = 0.7071 | cos(π/4) = 0.7071 |
| 9 AM | sin(3π/4) = 0.7071 | cos(3π/4) = -0.7071 |
| 2 PM | sin(7π/6) = -0.5 | cos(7π/6) = -0.866 |
| 8 PM | sin(5π/3) = -0.866 | cos(5π/3) = 0.5 |



\*Multiplying by the radius, 3.8197, would be acceptable but unnecessary for most tasks.