Microsoft: DAT210x Programming with Python for Data Science

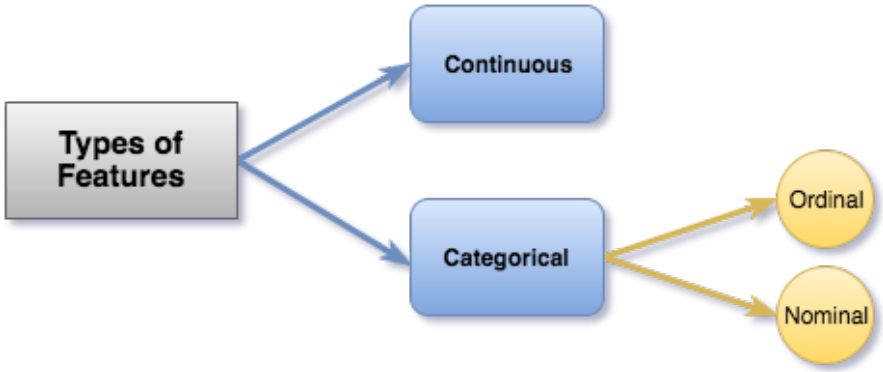
Feature Types

MOD7

There are many synonymous names for features you might hear them being referred to by, depending on the background of the speaker, as well as the context of the conversation:

- *Attribute* - Features are a quantitative attributes of the samples being observed
- *Axis* - Features are orthogonal axes of their *feature space*, if they are linearly independent
- *Column* - Features are represented as columns in your dataset
- *Dimension* - A dataset's features, grouped together can be treated as a *n-dimensional* coordinate space
- *Input* - Feature values are the input of data-driven, machine learning algorithms
- *Predictor* - Features used to predict other attributes are called predictors
- *View* - Each feature conveys a quantitative trait or perspective about the sample being observed
- *Independent Variable* - Autonomous features used to calculate others are like independent variables in algebraic equations

Although they have many names, any given feature will fall into one of two types:



Continuous Features

In the case of continuous features, there exist a measurable difference between possible feature values. Feature values usually are also a subset of all real numbers:

- Distance
- Time
- Cost
- Temperature

Categorical Features

With categorical features, there is a specified number of discrete, possible feature values. These values may or may not have ordering to them. If they do have a natural ordering, they are called ordinal categorical features. Otherwise if there is no intrinsic ordering, they are called nominal categorical features.

Nominal

- Car Models

- Colors
- TV Shows

Ordinal

- High-Medium-Low
- 1-10 Years Old, 11-20 Years Old, 30-40 Years Old
- Happy, Neutral, Sad