



## Microsoft: DAT210x Programming with Python for Data Science



Bookmarks

- ▶ Start Here
- ▶ 1. The Big Picture
- ▶ 2. Data And Features
- ▶ 3. Exploring Data
- ▶ 4. Transforming Data
- ▼ 5. Data Modeling

Lecture: Clustering  
Quiz



Lab: Clustering  
Lab



Lecture: Splitting Data  
Quiz



Lecture: K-Nearest  
Neighbors  
Quiz



Lab: K-Nearest Neighbors

## 5. Data Modeling &gt; Lecture: K-Nearest Neighbors &gt; Classification



Bookmark

## K-Nearest Neighbors

In this section, you're going to explore the K-Nearest Neighbors classifier. K-Neighbors and K-Means are similarly named, so people sometimes get the two confused, but they are actually different. K-Means is an unsupervised clustering algorithm, where K-Neighbors is a supervised classification algorithm. If clustering is the process of separating your samples into groups, then classification would be the process of assigning samples **into** those groups. Given a set of groups, take a set of samples and mark each sample as being a member of a group. Each group being the correct answer, label, or *classification* of the sample.

The K-Nearest Neighbors, or K-Neighbors classifier, is one of the simplest machine learning algorithms. Due to its simplicity, machine learning dabblers tend to start their journey by actually building out this classifier themselves. We urge you to do the same if you'd like to go that route.

The thought process behind K-Neighbors is that almost all information can be modeled on a continuous basis if you just "zoom in" close enough. We observe this in real life from the atomic to the cosmic scales. In atoms, you'll never encounter a proton resting in the clouds where electrons orbit. Instead you'll only find other electrons. Similarly, you won't encounter stray electrons hanging out in the nucleus where protons and electrons are packed. So if your purpose were to name the type of elementary particle and you knew its location, you could infer that its type based on that. If you look at a neighborhood on Zillow or RedFin and examined house prices, houses near one another tend to have similar sale values. Just by knowing the prices of neighboring houses, you have a good way of

Lab



discerning what an unidentified house might sell for\*. If you trained your eyes on the stars far away, those belonging to the same galaxy appear closer to one another than those belonging to different galaxies, and so on.

*\*Note: This example is actually an example of nearest neighbors regression instead of nearest neighbors classification, but it follows the same principle.*

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY  
OPENedX

