edX   **Microsoft:** DAT210x Programming with Python for Data Science

---

5. Data Modeling > Lecture: Regression > When To Use It

🔖 Bookmarks

🔖 **Bookmark**

▸ 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

When Should I Use Linear Regression?

Linear regression is widely used in all disciplines for forecasting upcoming feature values by extrapolating the regression line, and for estimating current feature values by interpolating the regression curve over existing data. It is an extremely well-understood and interpretable technique that run very fast, and produces reasonable results as long as you do not extrapolate too far away from your training data.

One of the main advantages of linear regression over other machine learning algorithms is that even though it's a supervised learning technique, it doesn't force you to fine tune a bunch of parameters to get it working. You can literally just dump your data into it and let it produce its results.

You can use linear regression if your features display a significant correlation. The stronger the feature correlation, it being closer to +1 or -1, the better and more accurate the linear regression model for your data will be. The questions linear regression helps you answer are which independent feature inputs relate to the dependent feature output variable, and the degree of that relationship.

In business, linear regression is often used to forecast sales. By finding a correlation between time and the number of sales, a company can predict their near-terms future revenue, which will then help them budget accordingly. Linear regression can also be used to assess risk. Before issuing a loan, most banks will consider many features or aspects about their customers, and run a regression to see if it is worthwhile for them to borrow the money, of if their return on investment is insignificant or even positive for that matter.

In the sciences, geologists train linear regression against historic records to calculate the rate of glacier snow melting, and can use it extrapolate how long it'll take for it to all disappear. Oil engineers do the same while calculating how much is potentially left. When measuring experimental results, chemists use linear regression to empirically calculate and validate concentrations and expected reactions. And of course there are many more uses.

POWERED BY
OPENedX