

Filter

- Logistic regression (35 min)
- Classification (70 min)

Data

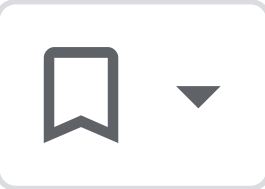
- Working with numerical data (85 min)
  - Introduction (3 min)
  - How a model ingests data with feature vectors (5 min)
  - First steps (5 min)
  - Programming exercises (10 min)
  - Normalization (20 min)
  - Binning (15 min)
  - Scrubbing (5 min)
  - Qualities of good numerical features (5 min)
  - Polynomial transforms (5 min)
  - Test your knowledge (10 min)



Home > Products > Machine Learning > ML Concepts > Crash Course

Was this helpful?

# Working with numerical data



Send feedback

Estimated module length: 85 minutes

## Learning objectives

- Understand feature vectors.
- Explore your dataset's potential features visually and mathematically.
- Identify outliers.
- Understand four different techniques to normalize numerical data.
- Understand binning and develop strategies for binning numerical data.
- Understand the characteristics of good continuous numerical features.

## Prerequisites:

This module assumes you are familiar with the concepts covered in the following module:

- [Introduction to Machine Learning](#)

ML practitioners spend far more time evaluating, cleaning, and transforming data than building models. Data is so important that this course devotes three entire units to the topic:

- Working with numerical data (this unit)
- [Working with categorical data](#)
- [Datasets, generalization, and overfitting](#)

This unit focuses on **numerical data**, meaning integers or floating-point values that behave like numbers. That is, they are additive, countable, ordered, and so on. The next unit focuses on **categorical data**, which can include numbers that behave like categories. The third unit focuses on how to prepare your data to ensure high-quality results when training and evaluating your model.

Examples of numerical data include:

- Temperature
- Weight
- The number of deer wintering in a nature preserve

In contrast, US postal codes, despite being five-digit or nine-digit numbers, don't behave like numbers or represent mathematical relationships. Postal code 40004 (in Nelson County, Kentucky) is not twice the quantity of postal code 20002 (in Washington, D.C.). These numbers represent categories, specifically geographic areas, and are considered categorical data.

## Key terms:

- [Categorical data](#)
- [Numerical data](#)

Help Center

[Previous](#)  
Test your knowledge (10 min)

[Next](#)  
How a model ingests data with feature vectors (5 min)

Was this helpful?



Send feedback

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](#), and code samples are licensed under the [Apache 2.0 License](#). For details, see the [Google Developers Site Policies](#). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2024-10-09 UTC.

Connect

Blog

Instagram

LinkedIn

X (Twitter)

YouTube

Programs

Google Developer Groups

Google Developer Experts

Accelerators

Women Techmakers

Google Cloud & NVIDIA

Developer consoles

Google API Console

Google Cloud Platform Console

Google Play Console

Firebase Console

Actions on Google Console

Cast SDK Developer Console

Chrome Web Store Dashboard

Google Home Developer Console

Google for Developers

Android

Chrome

Firebase

Google Cloud Platform

Google AI

All products

Terms | Privacy

Sign up for the Google for Developers newsletter

Subscribe

English