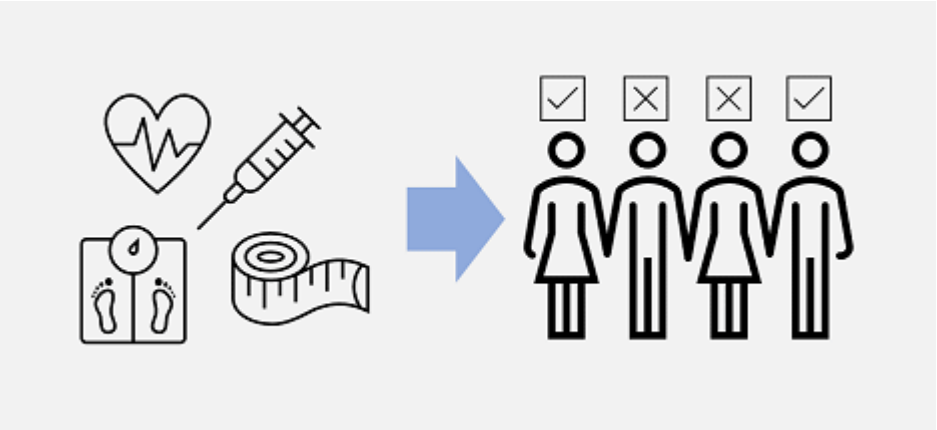
100 XP

Introduction

2 minutes

Classification is a form of machine learning in which you train a model to predict which category an item belongs to. For example, a health clinic might use diagnostic data such as a patient's height, weight, blood pressure, blood-glucose level to predict whether or not the patient is diabetic.



Categorical data has distinct 'classes', rather than numeric values. Some kinds of data can be either numeric or categorical: the time to run a race could be a time in seconds, or we could split times into classes of ‘fast’, ‘medium’ and ‘slow’ - categorical. While other kinds of data can only be categorical, such as a type of shape - ‘circle’, ‘triangle’, or ‘square’.

Prerequisites

- Knowledge of basic mathematics
- Some experience programming in Python

Learning objectives

In this module, you will:

- When to use classification
- How to train and evaluate a classification model using the Scikit-Learn framework

---

Next unit: What is classification?

Continue >

How are we doing? ☆ ☆ ☆ ☆ ☆