

DATA SCIENCE LEARNING PATH-1

What are machine learning models?

The model is the core component of machine learning, and ultimately what we're trying to build. A model might estimate how old a person is from a photo, predict what you might like to see on social media, or decide where a robotic arm should move. In our scenario, we want to build a model that can estimate the best boot size for a dog based on their harness size.

Models can be built in many ways. For example, a traditional model that simulates how an airplane flies is built by people, using knowledge of physics and engineering. Machine-learning models are special; rather than being edited by people so that they work well, machine learning models are shaped by data. They learn from experience.

How to think about models

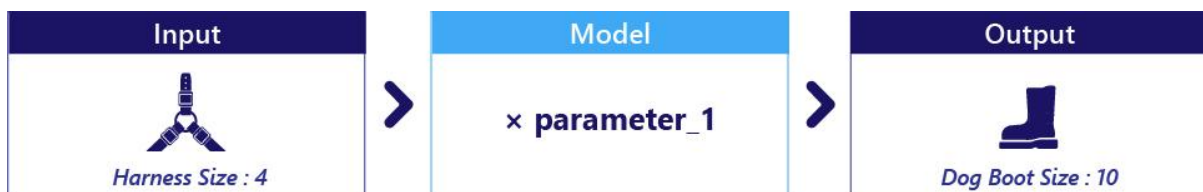
You can think of a model as a function that accepts data as an input and produces an output. More specifically, a model uses input data to estimate something else. For example, in our scenario, we want to build a model that's given a harness size and estimates boot size:



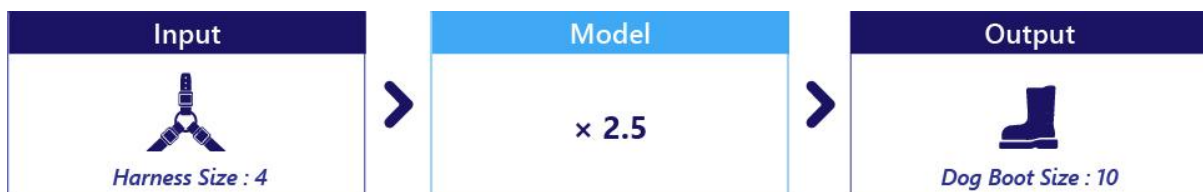
Harness size and dog boot size are data; they aren't part of the model. Harness size is our input, dog boot size is the output.

Models are often simple code

Models are often not meaningfully different from simple functions with which you're already familiar. Like other code, they contain logic and parameters. For example, the logic might be *multiply the harness size by parameter_1*:



If parameter_1 here was 2.5, our model would multiply harness size by 2.5 and return the result:

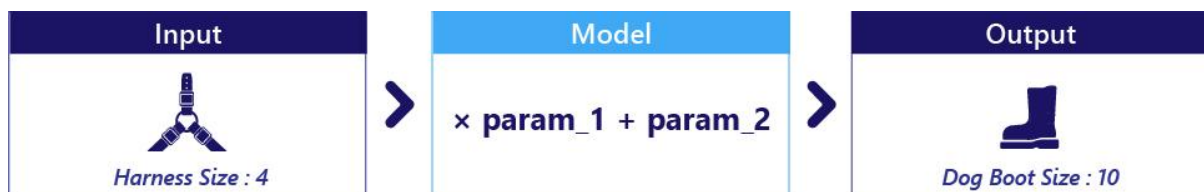


Select a model

There are many model types, some simple and some complex.

Like all code, simpler models are often the most reliable and easy to understand, while complex models can potentially perform impressive feats. Which kind of model you should choose depends on your goal. For example, medical scientists often work with models that are relatively simple, because they're reliable and intuitive. By contrast, AI-based robots typically rely on complex models.

The first step in machine learning is selecting the kind of model that you'd like to use. So, we're choosing a model based on its internal logic. For example, we might select a two-parameter model to estimate dog boot size from harness size:

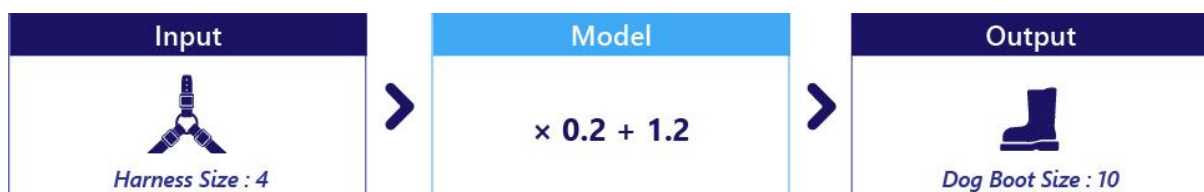


Notice how we selected a model based on how it works logically, but not based on its parameter values. In fact, at this point, the parameters haven't been set to any particular value.

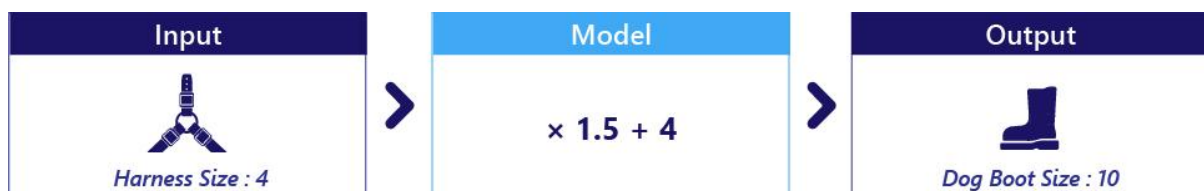
Parameters are discovered during training

The human designer doesn't select parameter values. Instead, parameter values are set to an initial guess, then adjusted during an automated learning process called training.

Given our selection of a two-parameter model, we start by providing random guesses for our parameters:



These random parameters mean the model isn't good at estimating boot size, so we perform training. During training, these parameters are automatically changed to two new values that give better results:



Exactly how this process works is something we progressively explain throughout your learning journey.