March 22, 2022

What is Machine Learning?

Updated: Jan 12, 2024

Machine learning is a breakthrough technology that can revolutionize the way we analyze and make decisions based on data. With the ability to adapt and learn, machine learning algorithms can handle vast amounts of data and provide accurate predictions and insights. This makes it an indispensable tool for optimizing complex data sets, like those found in bioinformatics, where it can uncover hidden patterns and interactions. The power and flexibility of machine learning surpass human-designed analytical tools and can lead to more efficient and effective decision-making. If you want to stay ahead of the curve in the data-driven world, machine learning is the way to go.

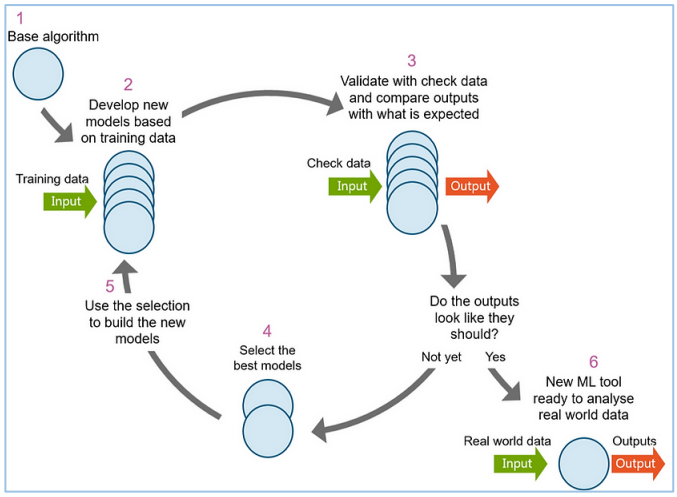
ML system has three parts: **data input**, **one or more hidden processing layers**, and **outputs**.

When we want to analyze a bunch of data, we first need to list it in a format that the computer can understand. This is **data input**. We then upload it into the machine for analysis.

The **outputs** we get from this analysis usually involve making some kind of prediction or categorization based on what the machine has learned. This allows us to answer questions like if a sample is cancerous or healthy, what signaling pathways are involved in a disease, or if changing an animal's diet will make it grow. It is important to know what kind of outputs you want before you start developing the system, as this helps the machine learn better.

The most exciting part of machine learning is the **hidden layer**. This is where the analysis model is found, which is the algorithm or set of calculations that turns data into outputs. To build this hidden layer, we start with a basic analysis framework and then feed in lots of training data. We teach the machine what we are looking for. There are many ways to do this, but usually, it is a cyclic process where the system is trained, tested, and trained again, allowing it to evolve and adapt until the outputs match what we expect. One example is

genetic algorithms, which use Darwinian selection to evolve the algorithm.**The general workflow for this evolution is described in the diagram below.**



1. Start by installing the software and create a basic algorithm to use as the analysis model.
2. Next, input some training data and let the system generate new analysis models.
3. Then check the models against a separate set of data to see how accurate they are.
4. Choose the best-performing models and get rid of the rest.
5. Use the selected models as a starting point, input more training data to generate more models.
6. Keep repeating this cycle until you are happy with the accuracy of the results and the machine learning system is ready to go!

Also, there are two main types of machine learning: supervised and unsupervised learning.

Supervised learning involves using labeled data, where you tell the machine what each data point represents. So, you can feed in different patient data sets labeled with cancer and those without. The machine learns to differentiate the two, which helps it detect cancer in new patients. With supervised learning, the machine can become better at differentiating sample types including non-human, which improves speed, accuracy, and early detection of cancer.

However, **unsupervised learning** uses unlabeled data, where only the measurements of interest are included without any indication of what they may be. This allows the machine to identify structures in the data and figure out what is important. Unsupervised learning is good for finding hidden patterns or clusters in the data and assessing how they relate to each other. For e.g., you can input data describing gene expression in cancer samples, and the machine will find which genes or groups of genes may be working together to cause the cancer.

Unsupervised learning is particularly suitable for analyzing large omics data sets to find previously unknown biological mechanisms. It can uncover unexpected or novel patterns in very complex data, which makes it exciting!

So now I hope you know more about machine Learning. There are more blog posts to come giving examples of applying this technology to interesting questions using biological big data. If you are eager to learn more about using Machine Learning in Biology, look at our video.

Written: Shelley Edmunds
Updated: Marie Sorivelle