What is Supervised and Unsupervised Learning?

Supervised Learning:

Suppose, you're doing a task under someone's supervision. So, someone is present there to judge whether you're getting the right answer or not. Similarly, in supervised learning, we have labels for all the data points and the model learns from that data.

Example:

Suppose, we have an email dataset and all the emails are labeled with either "Spam" or "Not Spam". So, each of our data points has one of the two labels "Spam" or "Not Spam".

Now when we get a new email, the job of the Machine Learning algorithm is to predict whether the email is Spam or Not Spam.

This is a Supervised Learning Problem.

Unsupervised Learning:

In unsupervised learning, we try to discover hidden patterns in data but we don't have any labels. The dataset here is a collection of examples without a specific desired outcome or correct answer.

Example:

In our email inbox, we have thousands of emails. And, say we want to read at any moment all the emails which are on the same topic. And we don't have any labels on the emails like Spam, Not Spam, etc.

So, we want to run a machine learning algorithm that groups together similar kinds of emails.

Now after running the algorithm, we find some natural groups of emails in the inbox, for example, emails about homework & classes, emails about hiking & adventures, etc.

So, this is an Unsupervised Learning problem where we don't have any labels on the dataset (i.e emails). But after running the algorithm we find some natural grouping of emails.

Some other examples of Unsupervised Learning are:

1. Google groups together all the articles that are about a similar story. (Google News)
2. Facebook trending stories are also just groupings of lots of individual articles and posts.

Challenges of Unsupervised Learning:

- Sometimes it's really expensive to get labels on the dataset.
- It's easier to understand how well a supervised learning algorithm is performing. We have to just check for example what percentage of spam and not spam emails are getting a label. But in the case of Unsupervised Learning, it's not that straightforward to check whether the emails are getting a good grouping by topic.