Machine Learning

Pawel Wocjan

University of Central Florida

Spring 2019

Sources for Slides

- I have extensively used the machine learning materials that have been prepared by Google.
  `https://developers.google.com/machine-learning/problem-framing/cases`
- Google has licensed these materials under the Creative Commons Attribution 3.0 License.
  `https://creativecommons.org/licenses/by/3.0/`

Outline

Common ML Problems

- In basic terms, ML is the process of training a piece of software, called a model, to make useful predictions using a data set.
- This predictive model can then serve up predictions about previously unseen data.
- We use these predictions to take action in a product; for example, the system predicts that a user will like a certain video, so the system recommends that video to the user.
- Often, people talk about ML as having two paradigms, supervised and unsupervised learning.
- However, it is more accurate to describe ML problems as falling along a spectrum of supervision between supervised and unsupervised learning.

Supervised Learning

- Supervised learning is a type of ML where the model is provided with labeled training data. But what does that mean?
- For example, suppose you are an amateur botanist determined to differentiate between two species of the Lilliputian plant genus (a completely made-up plant).
- The two species look pretty similar. Fortunately, a botanist has put together a data set of Lilliputian plants she found in the wild along with their species name.

Supervised Learning

- Here's a snippet of that data set:

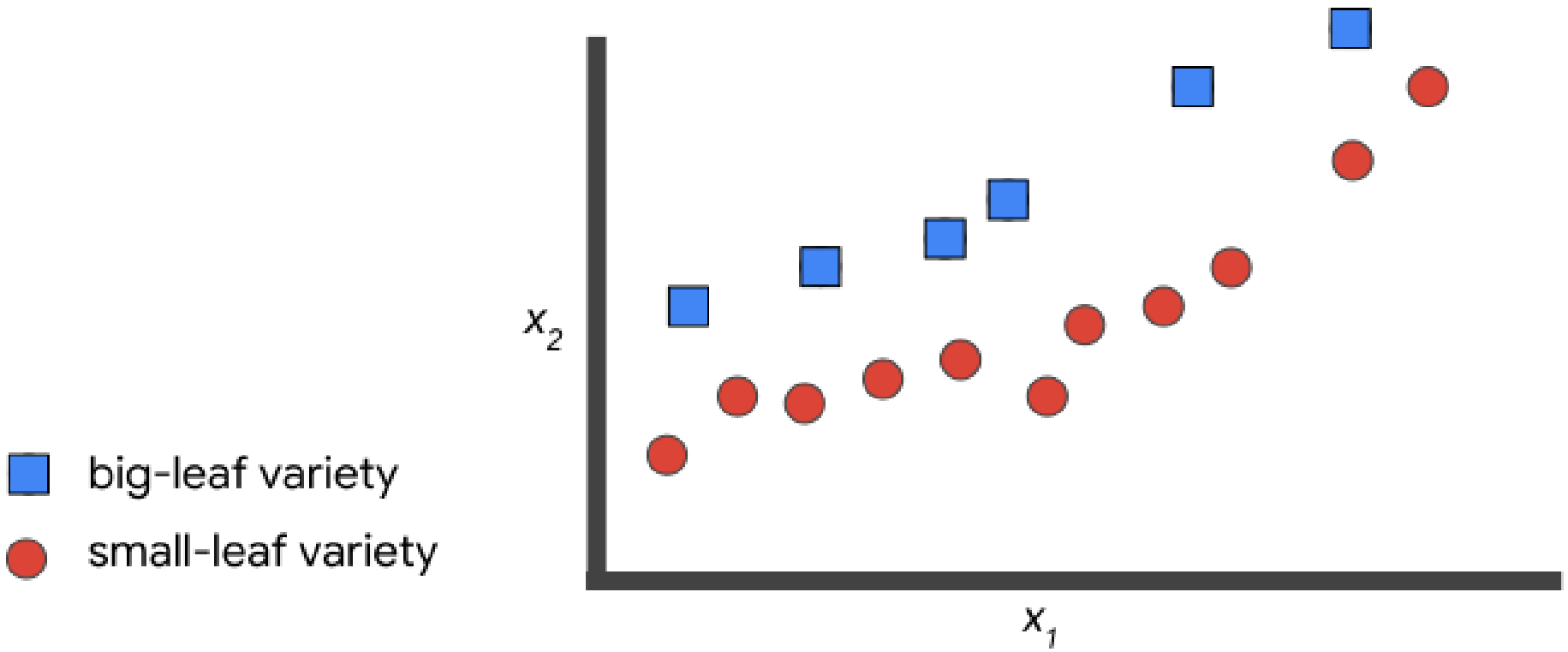
Leaf Width	Leaf Length	Species
2.7	4.9	small-leaf
3.2	5.5	big-leaf
2.9	5.1	small-leaf
3.4	6.8	big-leaf

Supervised Learning

- Leaf width and leaf length are the features, while the species is the label.
- A real life botanical data set would probably contain far more features (including descriptions of flowers, blooming times, arrangement of leaves) but still have only one label.
- Features are measurements or descriptions; the label is essentially the "answer."
- For example, the goal of the data set is to help other botanists answer the question, "Which species is this plant?"
- This data set consists of only four examples. A real life data set would likely contain vastly more examples.

Supervised Learning

- Suppose we graph the leaf width and leaf length and then color-code the species.

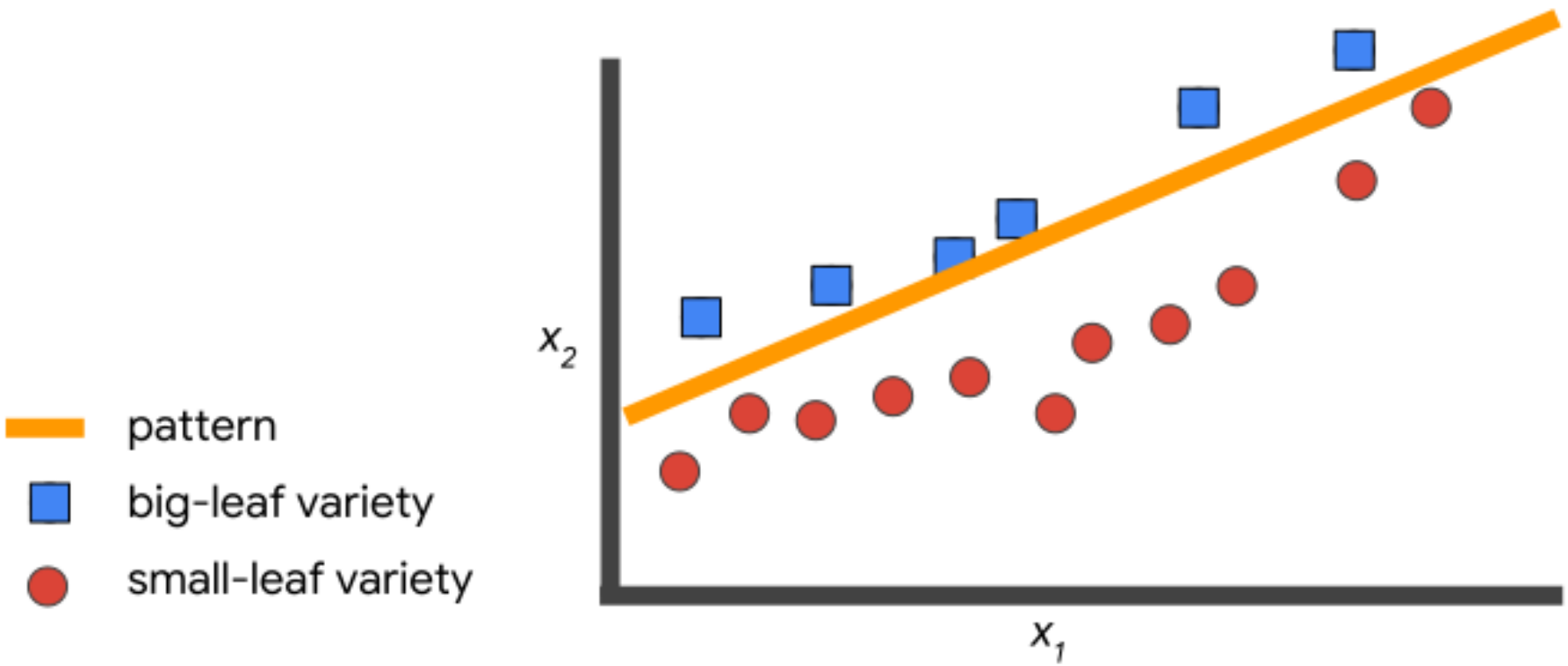


Supervised Learning

- In supervised machine learning, you feed the features and their corresponding labels into an algorithm in a process called training.
- During training, the algorithm gradually determines the relationship between features and their corresponding labels. This relationship is called the model.
- Often times in machine learning, the model is very complex.

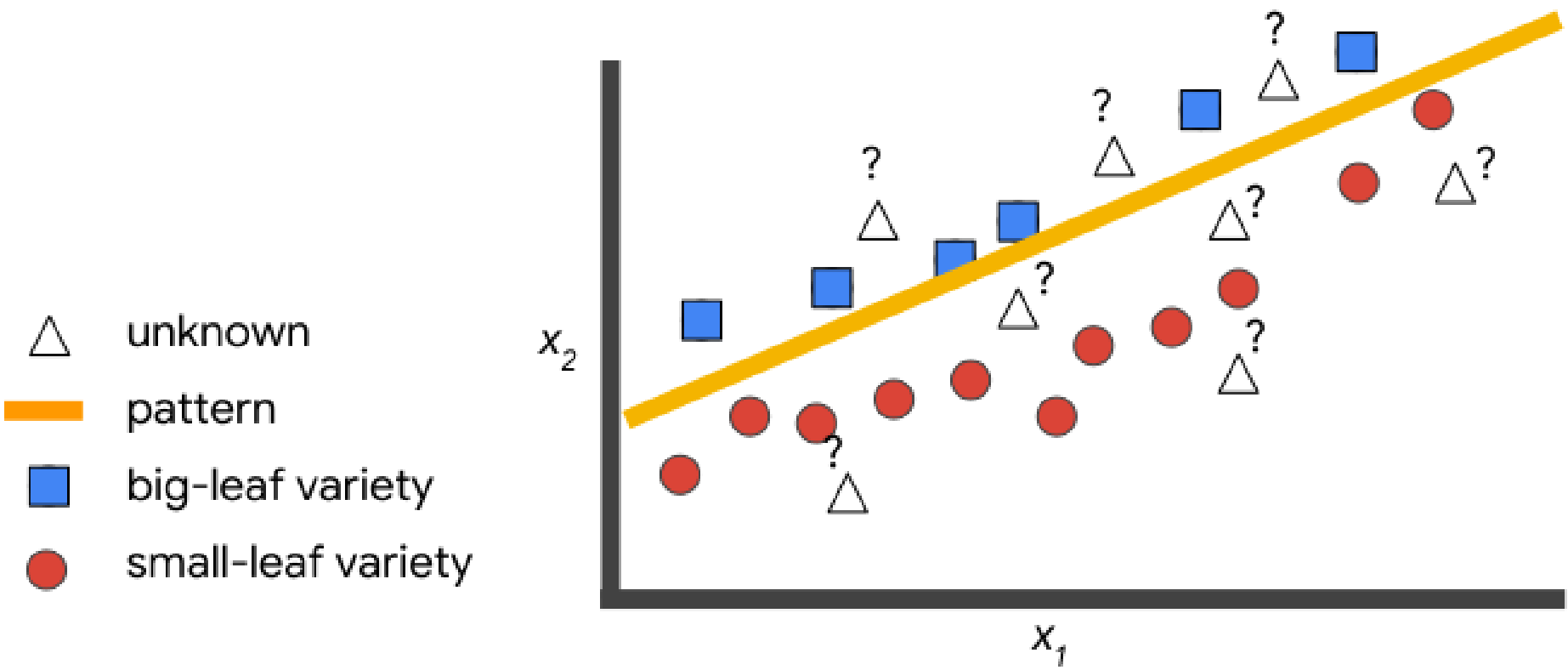
Supervised Learning

- However, suppose that this model can be represented as a line that separates big-leaf from small-leaf:



Supervised Learning

- Now that a model exists, you can use that model to classify new plants that you find in the jungle. For example:



Supervised Learning

- To tie it all together, supervised machine learning finds patterns between data and labels that can be expressed mathematically as functions.
- Given an input feature, you are telling the system what the expected output label is, thus you are supervising the training. The ML system will learn patterns on this labeled data.
- In the future, the ML system will use these patterns to make predictions on data that it did not see during training.

Supervised Learning

- An exciting real-world example of supervised learning is a study from Stanford University that used a model to detect skin cancer in images.

  `https://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/`

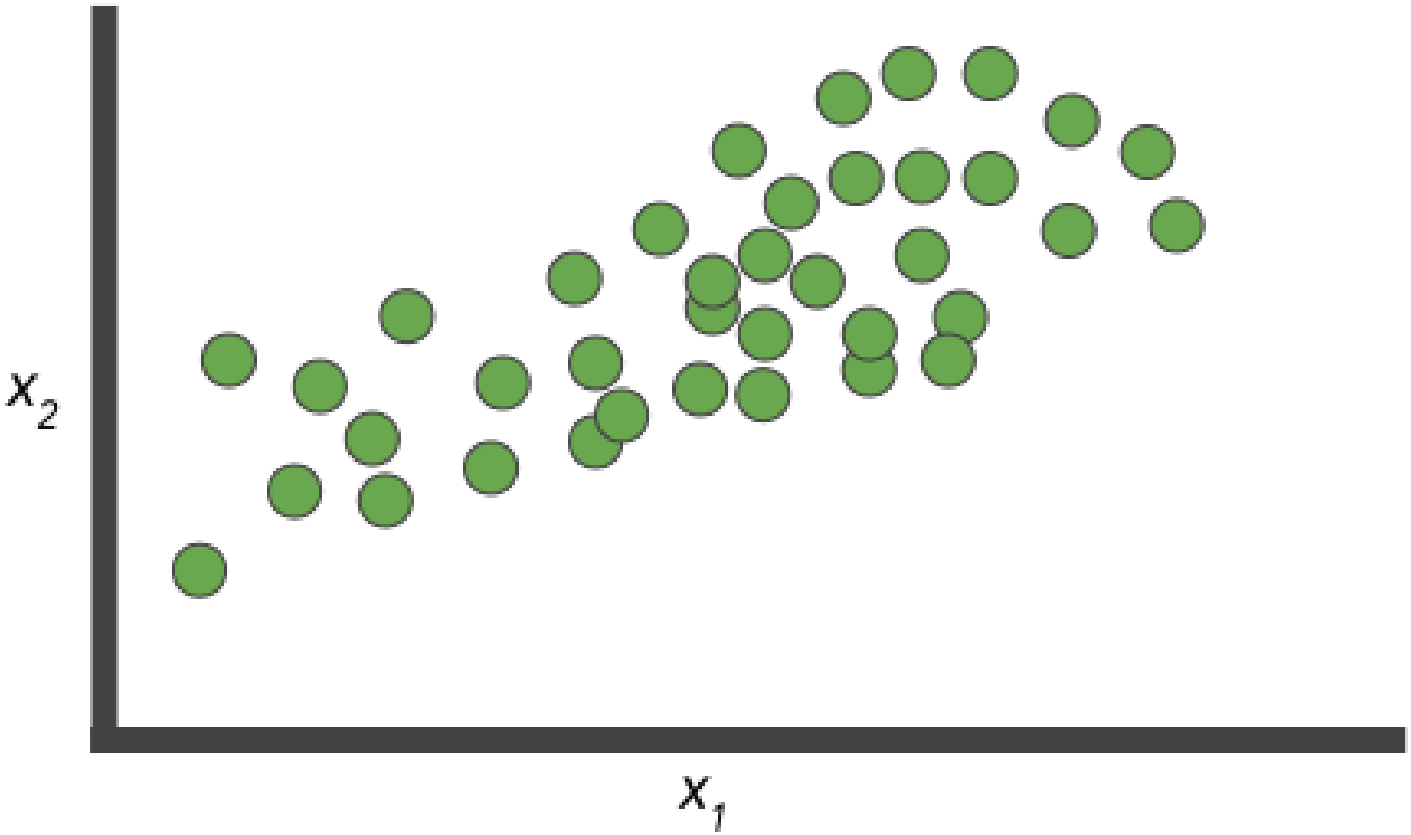
- In this case, the training set contained images of skin labeled by dermatologists as having one of several diseases. The ML system found signals that indicate each disease from its training set, and used those signals to make predictions on new, unlabeled images.

Unsupervised Learning

- In unsupervised learning, the goal is to identify meaningful patterns in the data.
- To accomplish this, the machine must learn from an unlabeled data set.
- In other words, the model has no hints how to categorize each piece of data and must infer its own rules for doing so.

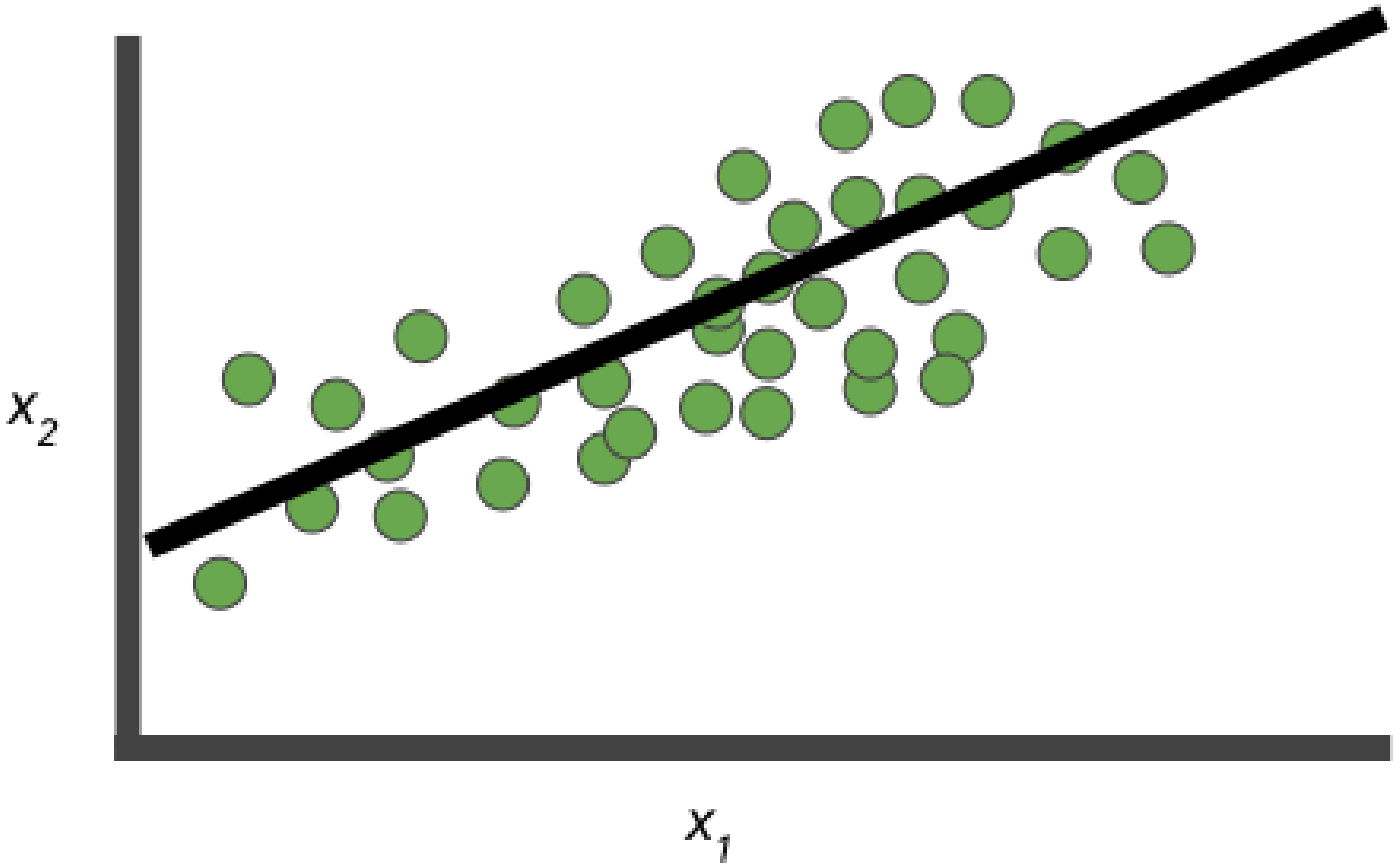
Unsupervised Learning

- All the examples are the same shape because we don't have labels to differentiate between examples of one type or another here:



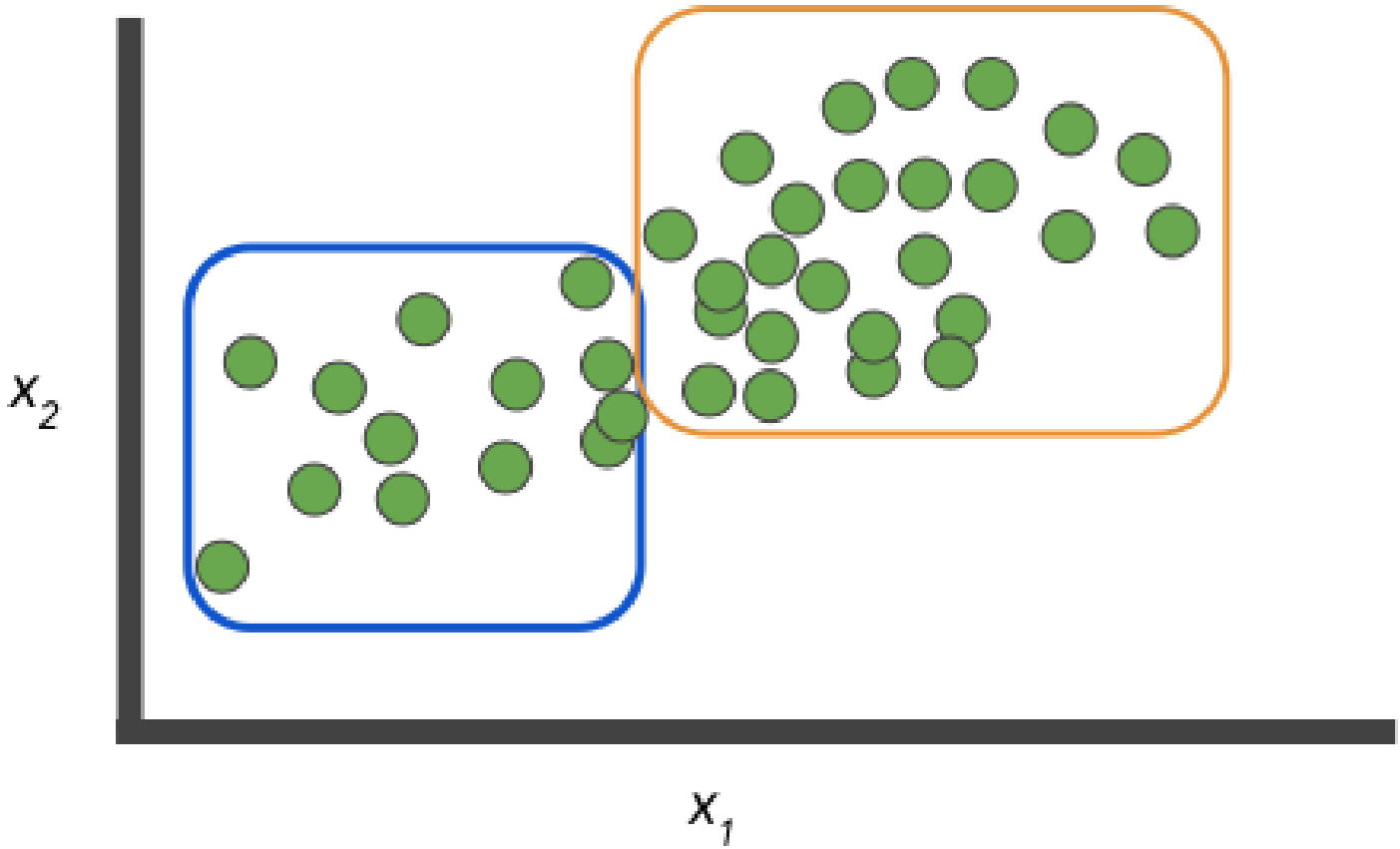
Unsupervised Learning

- Fitting a line to unlabeled points isn't helpful. We still end up with examples of the same shape on both sides of the line.



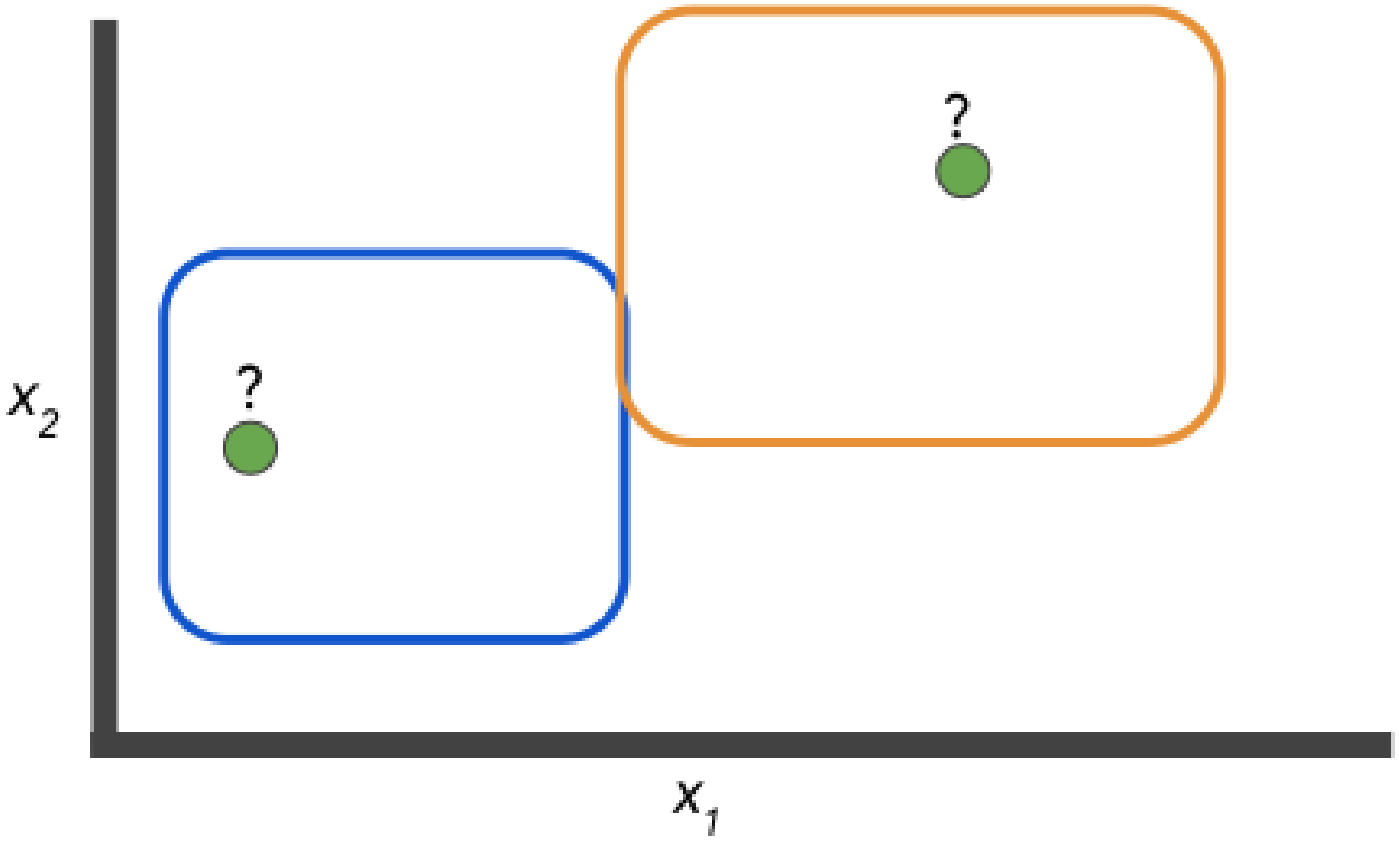
Unsupervised Learning

- Here, we have two clusters. What do these clusters represent? It can be difficult to say. Sometimes the model finds patterns in the data that you don't want it to learn, such as stereotypes or bias.



Unsupervised Learning

- However, when new data arrives, we can categorize it pretty easily, assuming it fits into a known cluster.



Unsupervised Learning

- But what if your photo clustering model has never seen a pangolin before?
  `https://en.wikipedia.org/wiki/Pangolin`
- Will the system cluster the new photo with armadillos or maybe hedgehogs?
  `https://en.wikipedia.org/wiki/Armadillo`
  `https://en.wikipedia.org/wiki/Hedgehog`
- This course will talk more about the difficulties of unlabeled data and clustering later on.

Reinforcement Learning

- An additional branch of machine learning is reinforcement learning (RL). Reinforcement learning differs from other types of machine learning.
- In RL you don't collect examples with labels. Imagine you want to teach a machine to play a very basic video game and never lose. You set up the model (often called an agent in RL) with the game, and you tell the model not to get a "game over" screen.
- During training, the agent receives a reward when it performs this task, which is called a reward function. With reinforcement learning, the agent can learn very quickly how to outperform humans.

Reinforcement Learning

- Designing a good reward function is difficult, and RL models are less stable and predictable than supervised approaches.
- You also need to provide a way for the agent to interact with the game to produce data, which means either building a physical agent that can interact with the real world or a virtual agent and a virtual world, either of which is a big challenge.
- Reinforcement learning is an active field of ML research, but in this course we'll not cover it. I cover RL in the course Advanced Artificial Intelligence.
- To learn more about RL, check out the following links:
  `https://gym.openai.com/`
  `http://incompleteideas.net/book/the-book-2nd.html`
  `https://www.alexirpan.com/2018/02/14/rl-hard.html`