**ML NOTES:**

* 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.
* For example,

Here, we have two clusters. (Note that the number of clusters is arbitrary). 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**](https://developers.google.com/machine-learning/glossary#bias_ethics).

* **Note: While it is very common, clustering is not the only type of unsupervised learning.**
* In **Reinforcement Learning** 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.
* The lack of a data requirement makes RL a tempting approach. However, designing a good reward function is difficult, and RL models are less stable and predictable than supervised approaches.

A screenshot of a cell phone

Description automatically generated

* Many machine learning systems produce models that encode knowledge and intelligence by interpreting signals differently than humans do. A neural network might interpret a word via an embedding, so "tree" is understood as something like, [0.37, 0.24, 0.2] and "car" as [0.1, 0.78, 0.9]. The neural network might use these representations to do accurate translations or sentiment analysis, but a human looking at the embeddings would find them very hard to understand.