Lecture 18
Unsupervised Learning

Dennis Sun
Stanford University
DATASCI / STATS 112

February 24, 2023

1. Supervised vs. Unsupervised Learning

2. Clustering

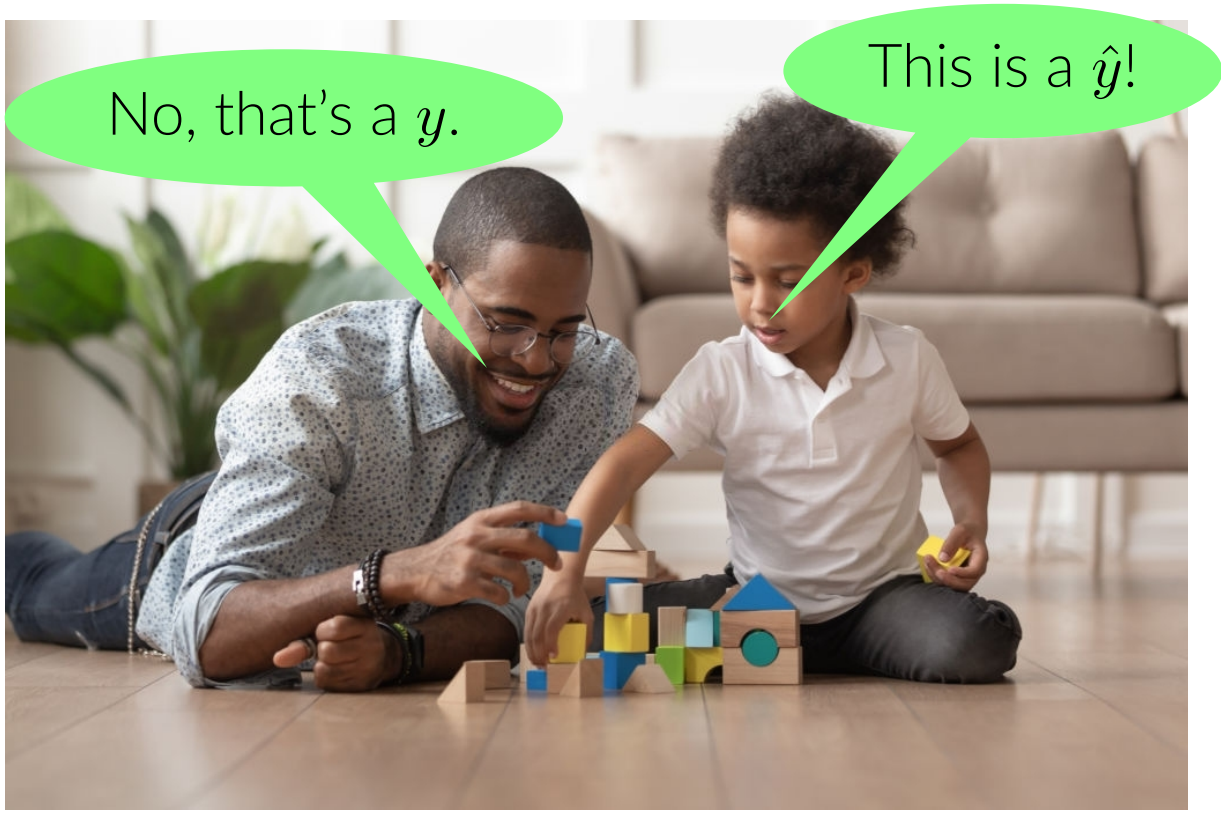
1 Supervised vs. Unsupervised Learning

2 Clustering

Supervised Learning

So far, in all the machine learning problems we've looked at, we had training data where the label y was known.

This is called **supervised learning**. Think of a child learning their shapes under the *supervision* of an adult.



Unsupervised Learning

What if we have training data without labels y?

This is called **unsupervised learning**. It is like a child who has been left to learn about shapes on their own without an adult to guide them.

1 Supervised vs. Unsupervised Learning

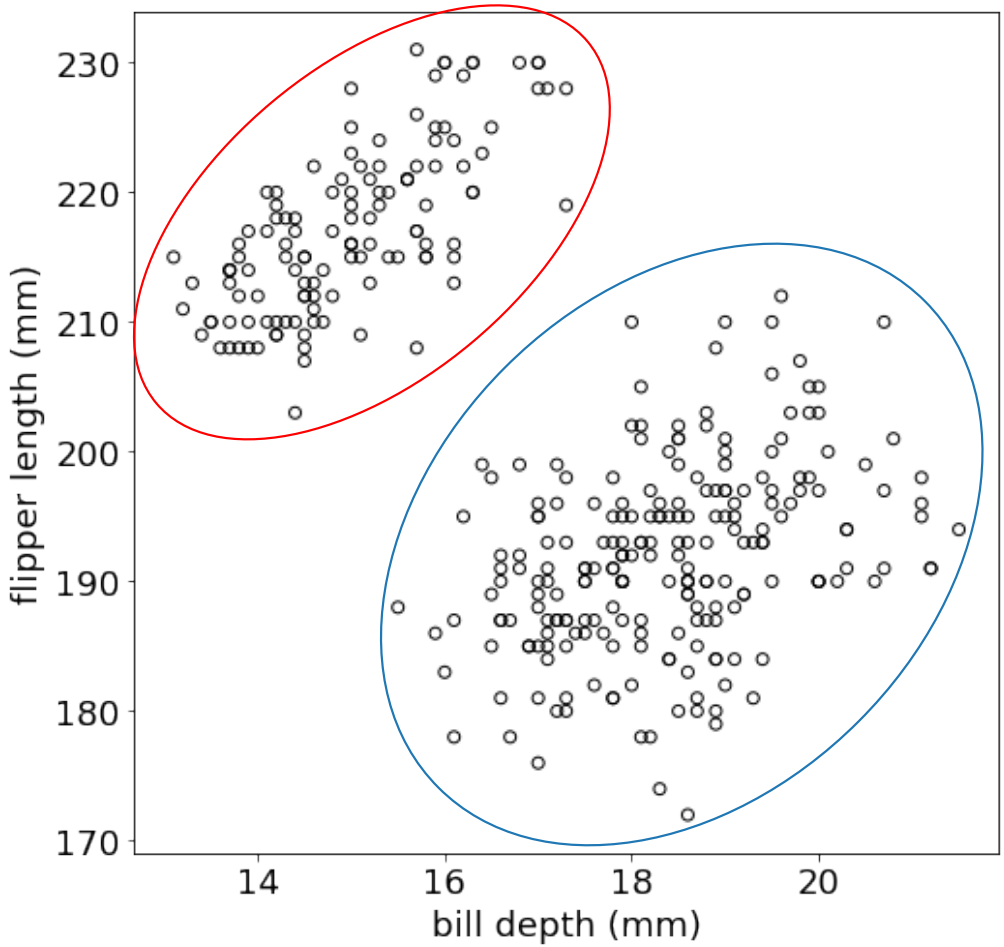
2 Clustering

Clustering

Clustering is one type of unsupervised learning.

The goal is to find groups of similar observations.

For example, consider the Palmer penguins data set.



There appear to be at least two distinct types of penguins.

Applications of Clustering

- **Ecology:** An ecologist wants to group organisms into types.
- **Market Segmentation:** A business wants to group their customers into types.
- **Language:** A linguist might want to identify different uses of ambiguous words like “set” or “run”.

In Assignment 6, you will explore another application of clustering to image compression!