Week 1: Part-time Data Science

Intro to Machine Learning

Dami Lasisi

Categories of ML

Supervised Learning (classification and regression)

- Predicts an outcome based on input data (features)
- Generalizes
- Requires already established data on the element we want to predict (target)

Unsupervised Learning (clustering and dimensionality reduction)

- Extracts structure from data
- Represents
- Does not require already established data on the element we want to predict

Supervised Learning

- Train a machine learning model to identify the relationships between the features and the target
- Make predictions on the target using the new feature data and the model that has been trained
- Primary goal: Train a model that can be generalized

Supervised Learning: Example

Features:

- ➡ # of bedrooms
- ➡ # of bathrooms
- ➡ # rooms
- ➡ garage
- ➡ zip-code
- ➡ sqft
- ➡ swimming pool
- ➡ age of neighbors
- ➡ age of house
- ➡ upgrades

Target:

- ➡ Price of House (e.g. $180,000)

Categories of Supervised Learning

- **Regression:**
  - Outcome (target) variable is continuous

- **Classification:**
  - Outcome (target) variable is binary or categorical
  - With the housing example, target variable will be price level (e.g. high, average, low)

Unsupervised Learning: Example

Features:

- ➡ # of bedrooms
- ➡ # of bathrooms
- ➡ # rooms
- ➡ garage
- ➡ zip-code
- ➡ sqft
- ➡ swimming pool
- ➡ age of neighbors
- ➡ age of house
- ➡ upgrades

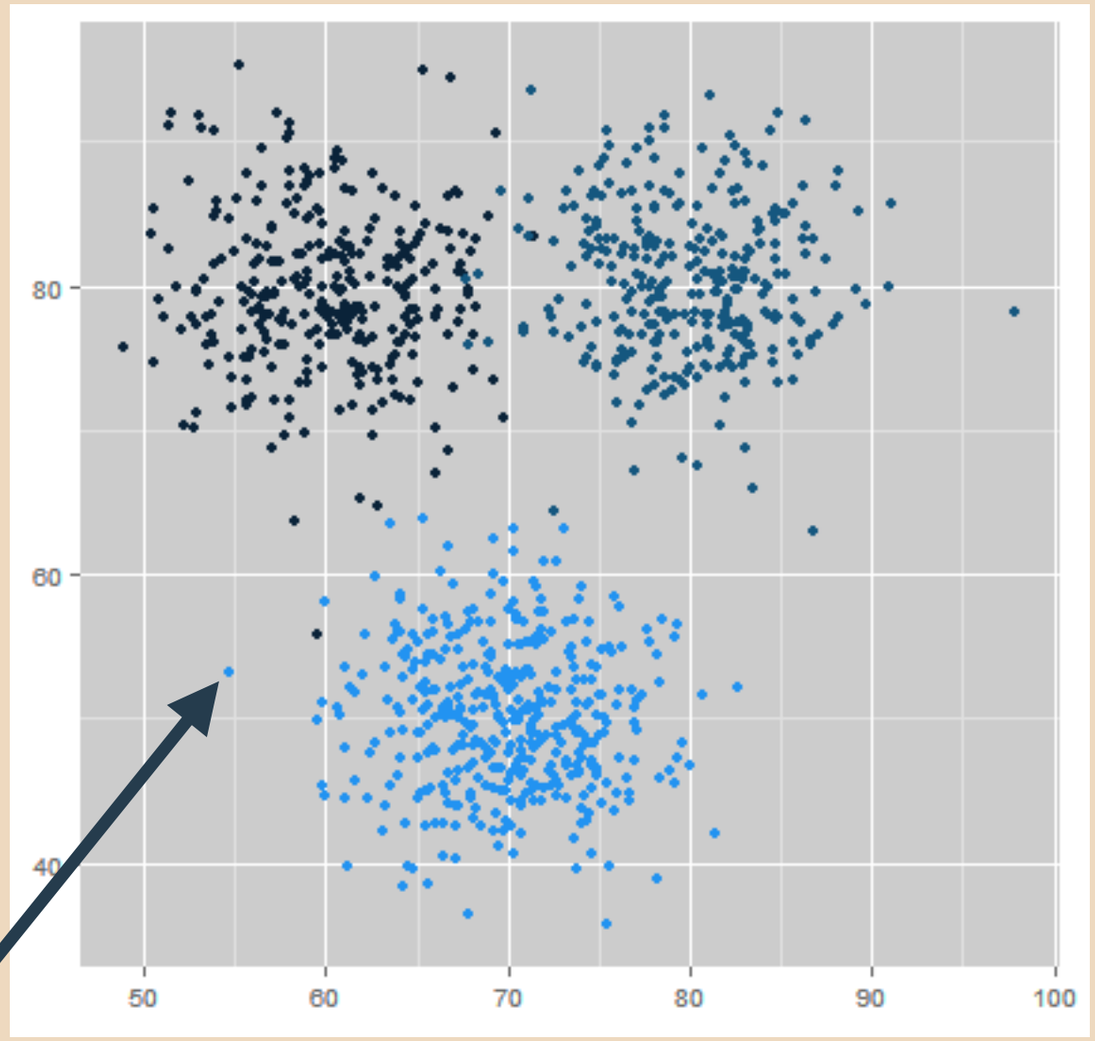
Target:

- ➡ Price of House

Common Types of Unsupervised Learning

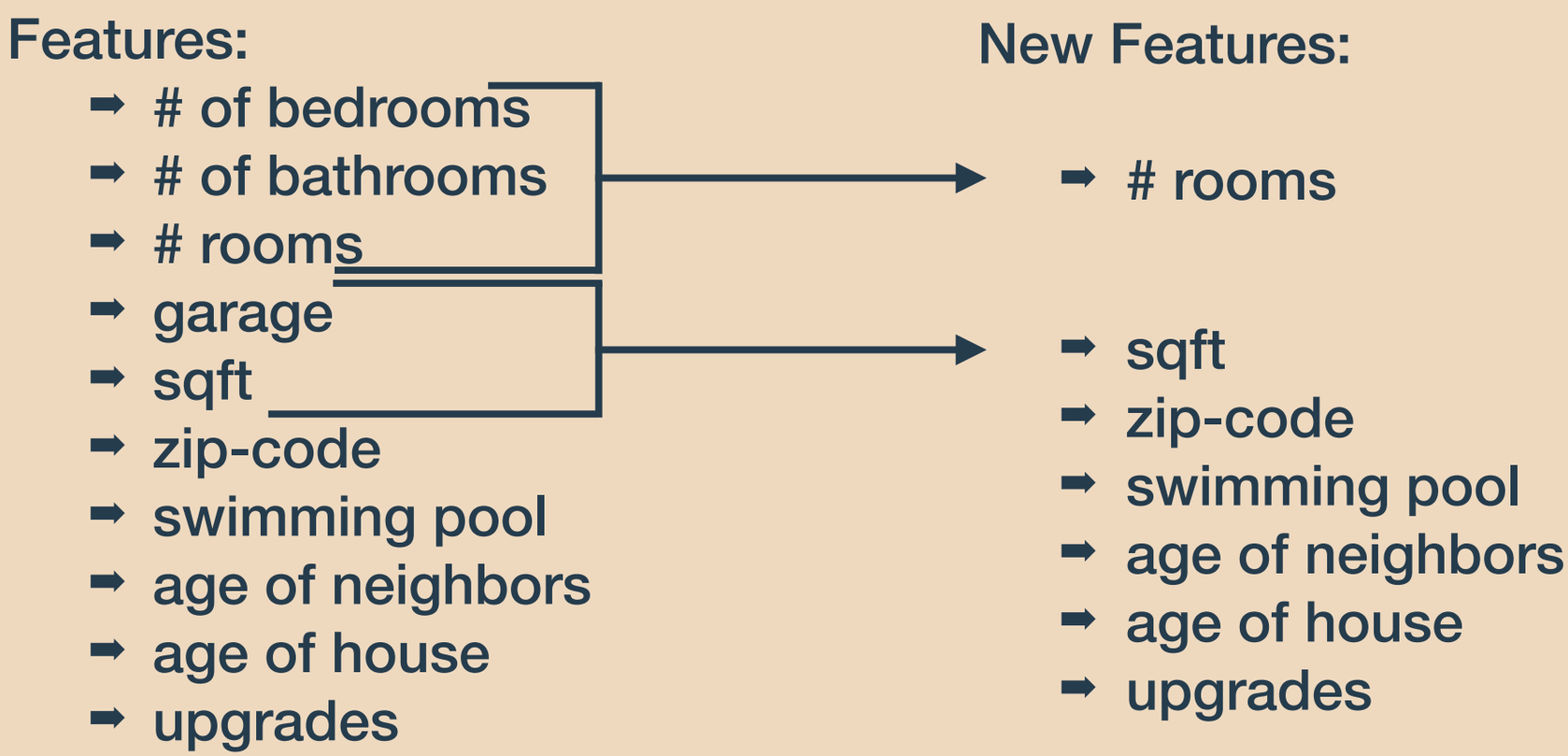
- **Clustering:**
  - Groups similar data points together
  - In theory, data points in the same group should have similar properties
- **Dimensionality Reduction:**
  - Extracts the features that captures the most variance in the data

Clustering: Example



One house

Dimensionality Reduction: Example



Algorithms

Help identify trends and relationships, explain the overall variance of the data

Features:

- ➡ # of bedrooms
- ➡ # of bathrooms
- ➡ # rooms
- ➡ garage
- ➡ zip-code
- ➡ sqft
- ➡ swimming pool
- ➡ age of neighbors
- ➡ age of house
- ➡ upgrades

$y = mx + b$

sqft(x) = 2,500

Price (y) = $285,000

$m = 114$, $b = 0$

Final Algorithm: Price(y) = $114x$