01_machine_learning_intro

February 28, 2024

1 What is Machine Learning, and how does it work?

Lesson 1 from Introduction to Machine Learning with scikit-learn

1.1 Agenda

- What is Machine Learning?
- What are the two main categories of Machine Learning?
- What are some examples of Machine Learning?
- How does Machine Learning "work"?

1.2 What is Machine Learning?

One definition: "Machine Learning is the semi-automated extraction of knowledge from data"

- **Knowledge from data**: Starts with a question that might be answerable using data
- **Automated extraction**: A computer provides the insight
- **Semi-automated**: Requires many smart decisions by a human

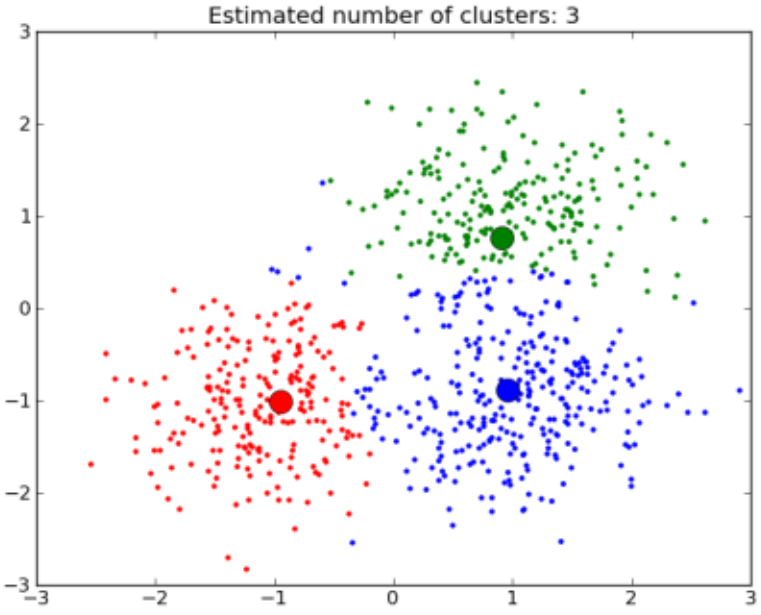
1.3 What are the two main categories of Machine Learning?

Supervised learning: Making predictions using data

- Example: Is a given email "spam" or "ham"?
- There is an outcome we are trying to predict

Unsupervised learning: Extracting structure from data

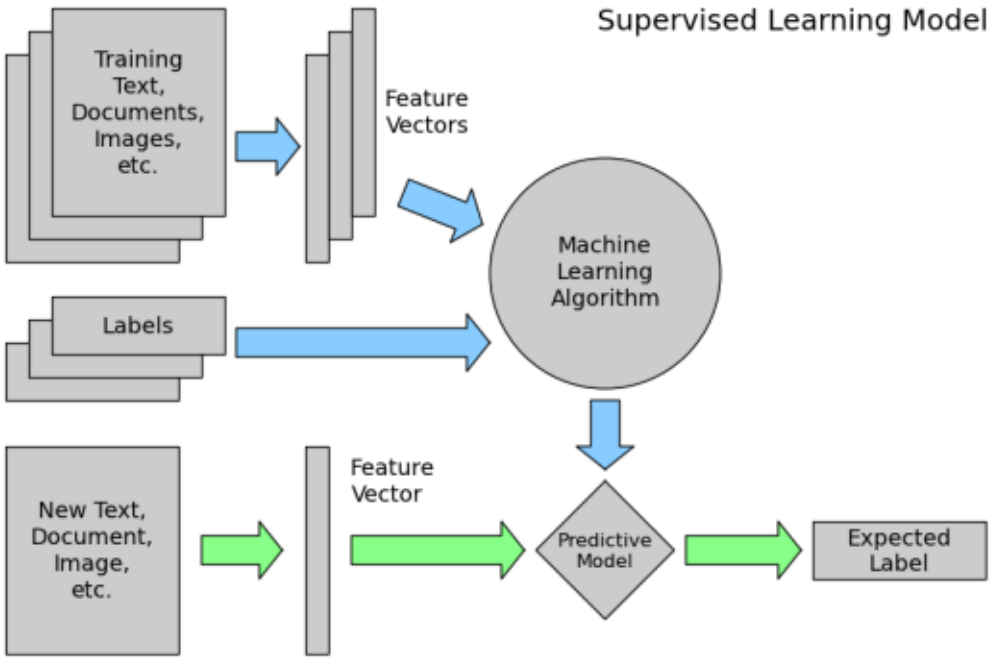
- Example: Segment grocery store shoppers into clusters that exhibit similar behaviors
- There is no "right answer"



1.4 How does Machine Learning "work"?

High-level steps of supervised learning:

1. First, train a **Machine Learning model** using **labeled data**
    - "Labeled data" has been labeled with the outcome
    - "Machine Learning model" learns the relationship between the attributes of the data and its outcome
2. Then, make **predictions** on **new data** for which the label is unknown



The primary goal of supervised learning is to build a model that "generalizes": It accurately predicts the **future** rather than the **past**!

1.5 Questions about Machine Learning

- How do I choose **which attributes** of my data to include in the model?
- How do I choose **which model** to use?
- How do I **optimize** this model for best performance?
- How do I ensure that I'm building a model that will **generalize** to unseen data?
- Can I **estimate** how well my model is likely to perform on unseen data?

1.6 Resources

- Book: An Introduction to Statistical Learning (section 2.1, 14 pages)
- Video: Learning Paradigms (13 minutes, starting at 36:02)

1.7 Comments or Questions?

- Email: kevin@dataschool.io
- Website: https://www.dataschool.io
- Twitter: @justmarkham