

Sketching out the Machine Learning Pipeline

Forrest Davis

August 29, 2025

Fall 2025 COSC410A: Applied Machine Learning
Colgate University

4 Key Parts of ML Pipeline

1. Data (samples and features)
2. Model
3. Objective Function
4. Optimization

Pipeline

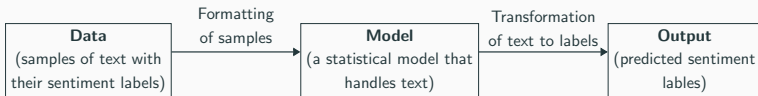
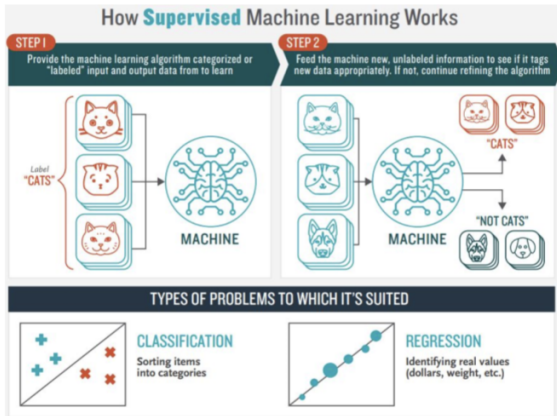
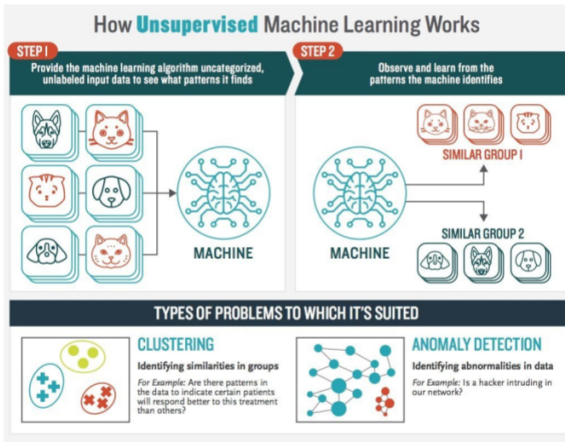


Figure 1: A sketch of the machine learning pipeline. The boxes represent elements where you, as an engineer, interject in the process. You must find, inspect, treat, and format the data, specify a model (and its computations), and inspect, validate, and evaluate the output of the model.

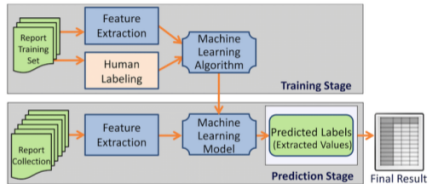
Supervised Learning



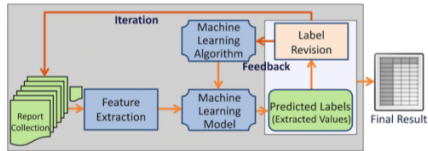
Unsupervised Learning



Offline vs Online Learning



(a) Traditional Batch Machine Learning



(b) Online Machine Learning

Limitations

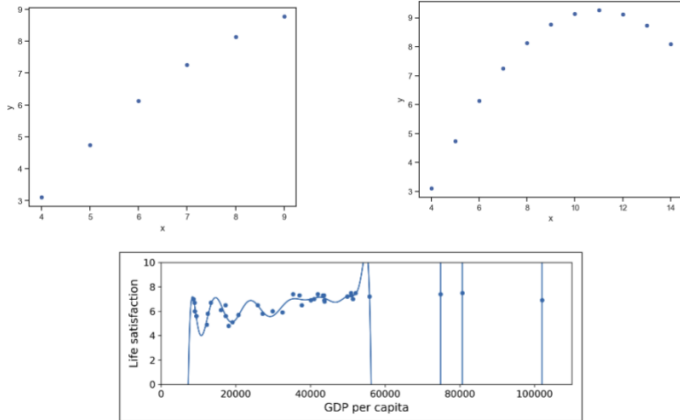


Figure 1-22. Overfitting the training data