Applied Machine Learning

Week 1

Instructor: Mariia Sidulova

- Work Experience
  - Medtronic Inc – Senior GenAI Data Scientist
  - Food and Drug Administration – Staff Scientist

- Education
  - University of Minnesota – Twin Cities
    - Bachelor of Biomedical Engineering
  - George Washington University
    - PhD in Biomedical Engineering

- Personal
  - I love to play tennis, snowboard and go to art galleries

Overview of Course

This course covers the most popular machine learning algorithms and the Python packages needed to implement them.

- You will learn supervised and unsupervised learning algorithms.
- Regression and classification problems.
- Linear and logistic regression, neural networks, decision trees, GenAI models.

Syllabus Highlights

Textbook

Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow by Aurelien Geron.

Assessments

- Homework (4 total): 40%
- Exams (2 total): 30% (15% each)
- Final Project: 25%
- Participation & Attendance: 5%

Homework Assignments – 40% of your grade

There will be 4 programming assignments.

You are given 2 weeks for each assignment.

Exams – 30% of your grade

There will be 2 exams. The first in week 5 and the second in week 10.

There is no final exam.

Class Participation – 5% of your grade

You deserve something for showing up for a t 6:15pm on Monday nights

Office Hours and Other Resources

Instructor

- Thursdays 6:30 – 7:30 and by appointment
- Virtually

What are you hoping to learn from this class?

What are we covering today?

- What is Machine Learning, Deep Learning, Artificial Intelligence?
- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Regression
- Anaconda, Github

AI vs ML vs DL

Artificial Intelligence (AI) ~ 1956

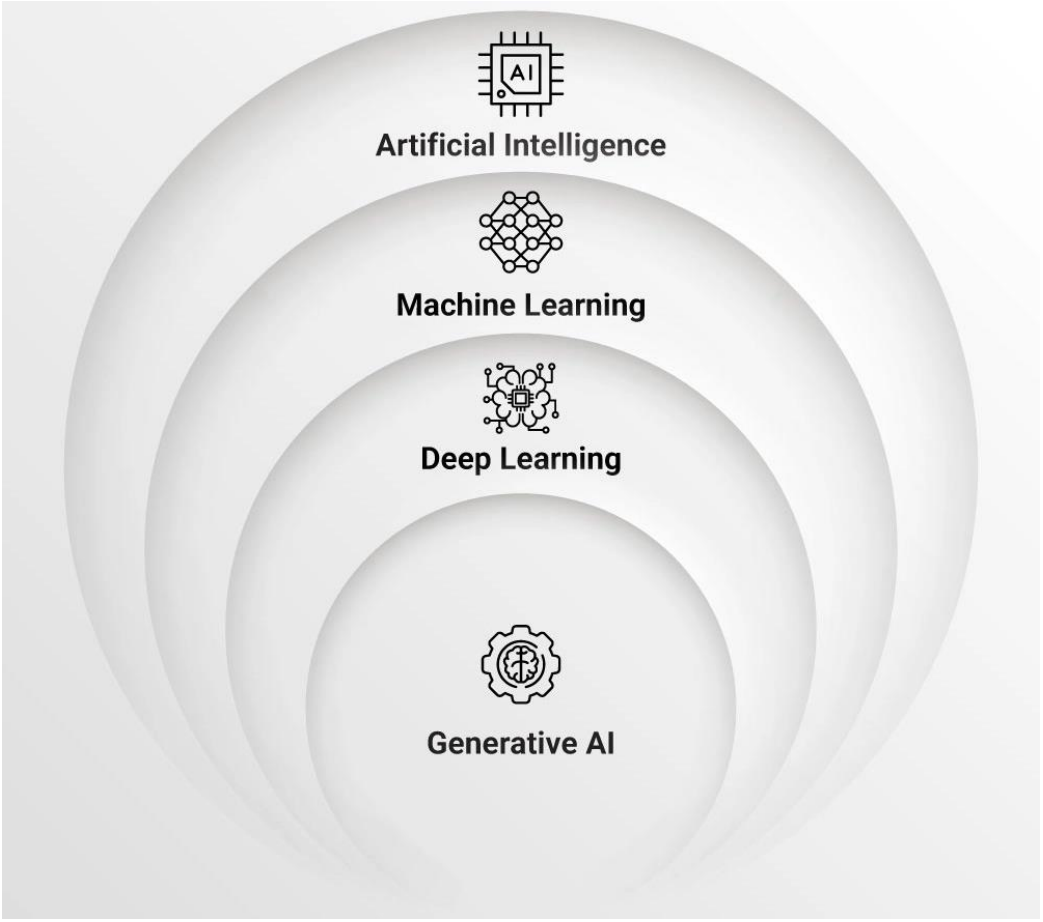
- Broad field aiming to create systems that can simulate human intelligence (reasoning, planning, problem-solving).

Machine Learning (ML) ~1959

- A subfield of AI that enables machines to learn from data and improve performance without being explicitly programmed.

Deep Learning (DL) ~ 1980s–2010s

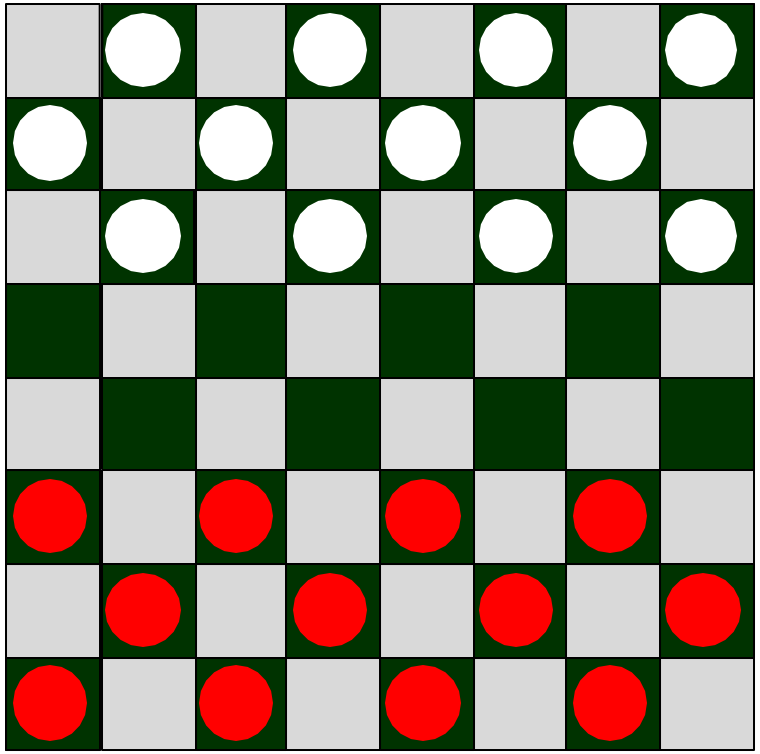
- The term "deep learning" became widely used in the to describe neural networks with many hidden layers that excel in vision, speech, and language tasks.



Machine Learning

"Field of study that gives computers the ability to learn without being explicitly programmed."

Arthur Samuel (1959)



Samuel wrote a research paper verifying that a computer can play a game of checkers better than the person who wrote the program. He is regarded as the person to coin the field machine learning that we know today.

Filtering system for spam email traditional

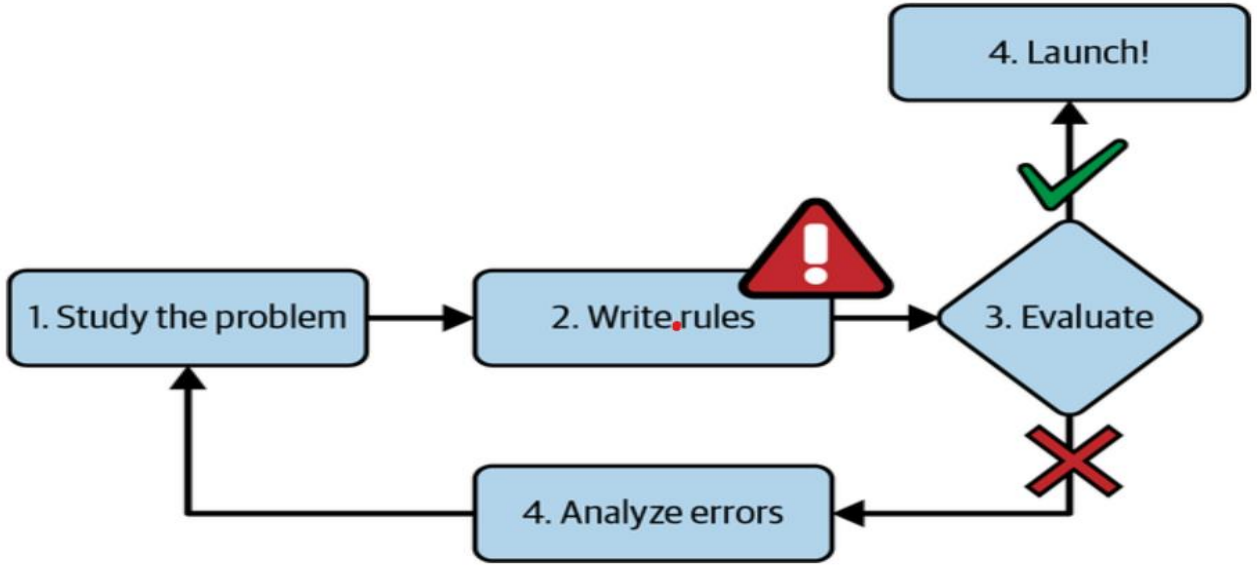


Figure 1-1. The traditional approach

The rules would get very long as spammers come up with more and more ways to bypass the rules.

Filtering system for spam email with ML

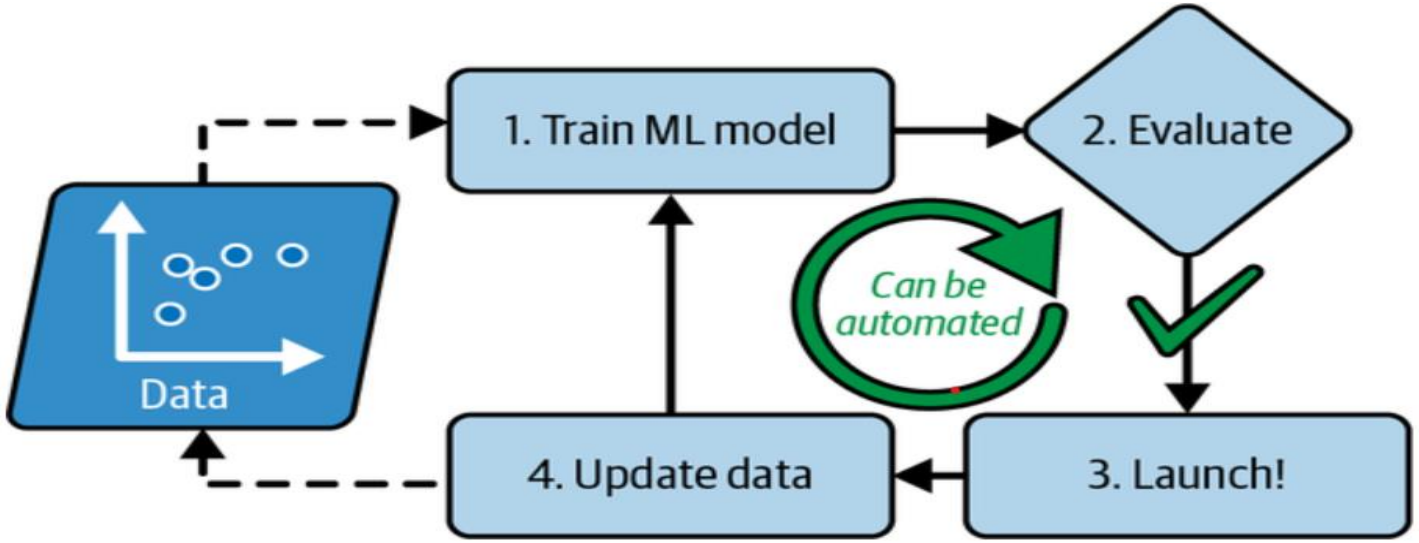


Figure 1-3. Automatically adapting to change

Machine learning uses an algorithm that finds patterns in the input data that indicates spam. The algorithm doesn't change as the data does.

When ML is needed?

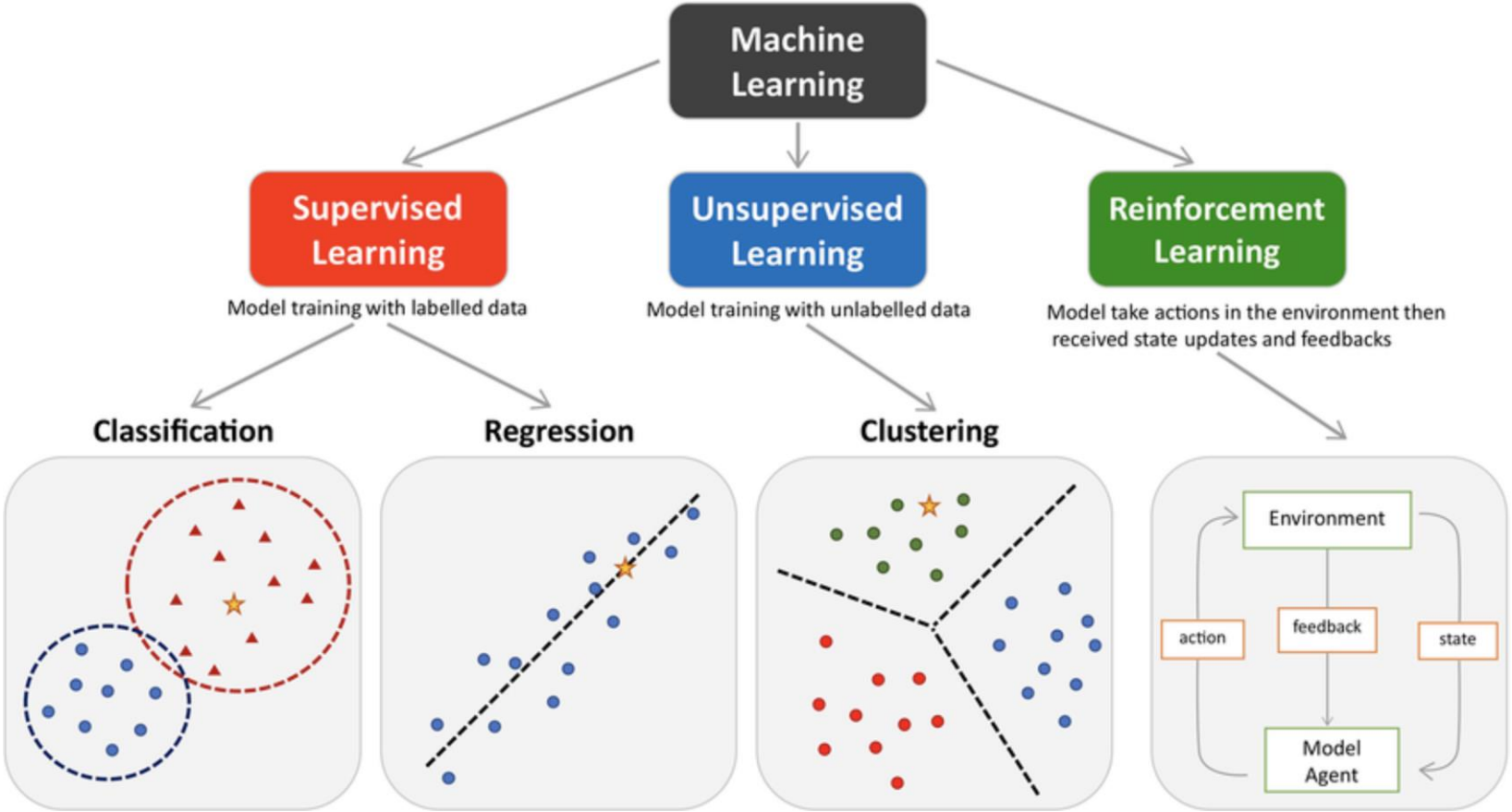
- Simplifies problems that need many rules or fine-tuning
- Finds solutions where traditional methods fail
- Adapts easily to changing environments
- Extracts insights from large, complex data

Where ML is used?

Where ML is used?

- Image Classification – look for defects in a production line, detect tumors in a brain scan, detects friends in a social media post.
- Natural Language Processing – chatbots, summarize content, flag offensive comments on a blog, classify content like news articles or movies.
- Predictions – forecast earnings, predict stock prices.
- Fraud detection – flag suspicious credit card activity.
- Voice recognition - Siri
- Recommendation Systems – Netflix suggest shows you may like.
- And more!

Types of ML


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Model training with labelled data
Model training with unlabelled data
Model take actions in the environment then received state updates and feedbacks
Classification
Regression
Clustering
Environment
action
feedback
state
Model Agent

Supervised Learning

- Supervised learning uses “labeled” training data to help make predictions on test data.

Input (features):
Home size, number of bedrooms, lot size etc.

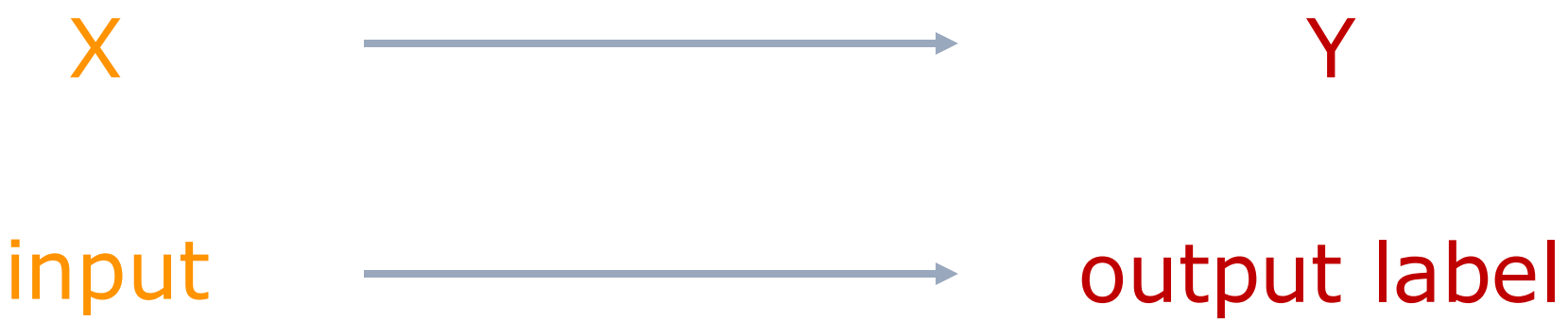
ML Algorithm:
Makes prediction

$

Output (target):
Home Price

Supervised machine learning uses data with known inputs and outputs to train an algorithm. The trained algorithm can then make predictions on future inputs where the output is not known yet.

Supervised Learning



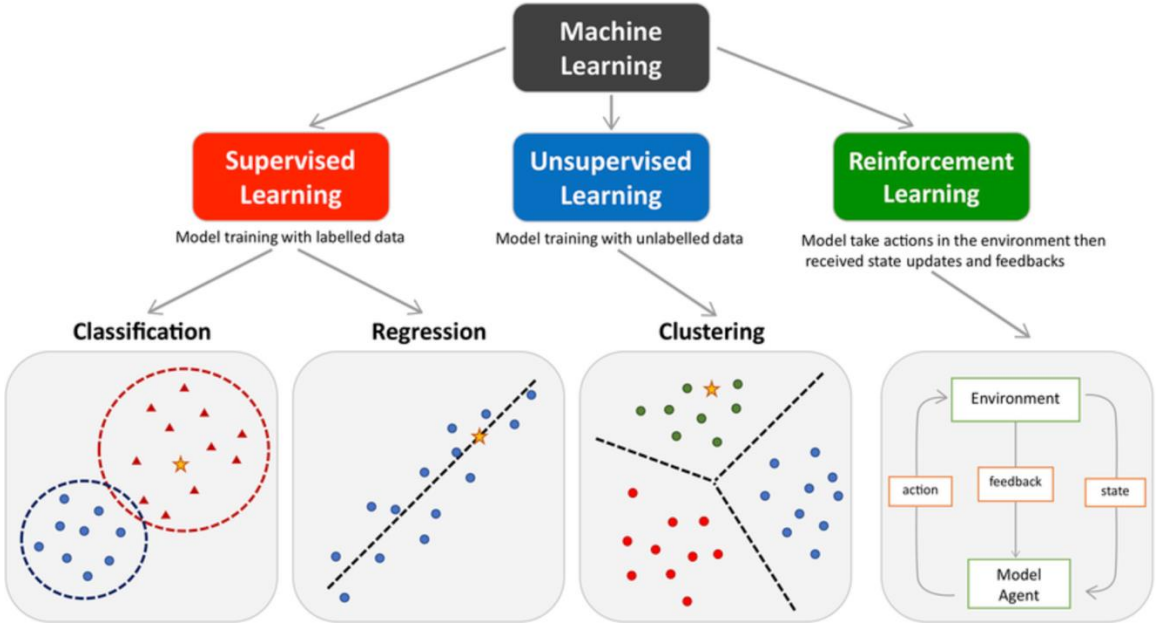
Learns from being given "right answers"

Input (X)	Label (Y)	Application
email	spam? (0/1)	spam filtering
audio	text transcripts	speech recognition
English	Spanish	machine translation
ad, user info	click? (0/1)	online advertising
image, radar info	position of other cars	self-driving car
image of phone	defect? (0/1)	visual inspection

Supervised Learning: Regression vs Classification Problems

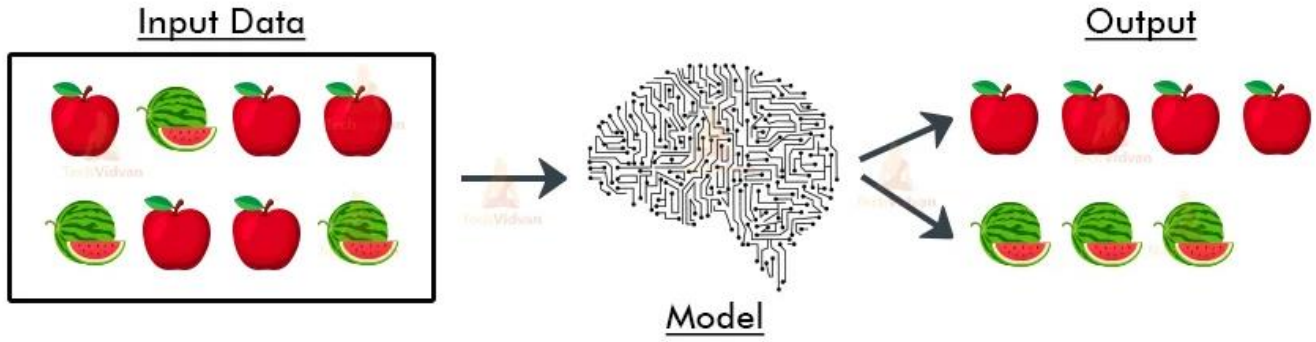
Output of a regression problems are continuous numeric values – example home price

Output of classification problems are discrete – for example classification of cats, dogs

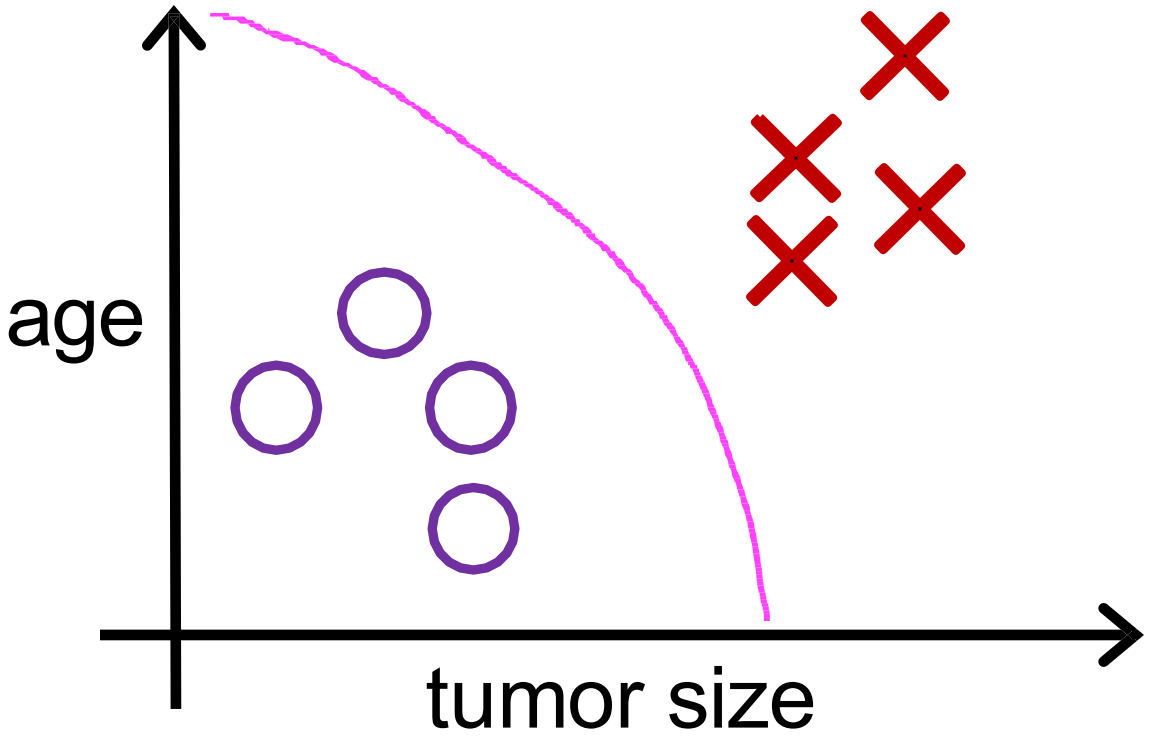


Unsupervised learning

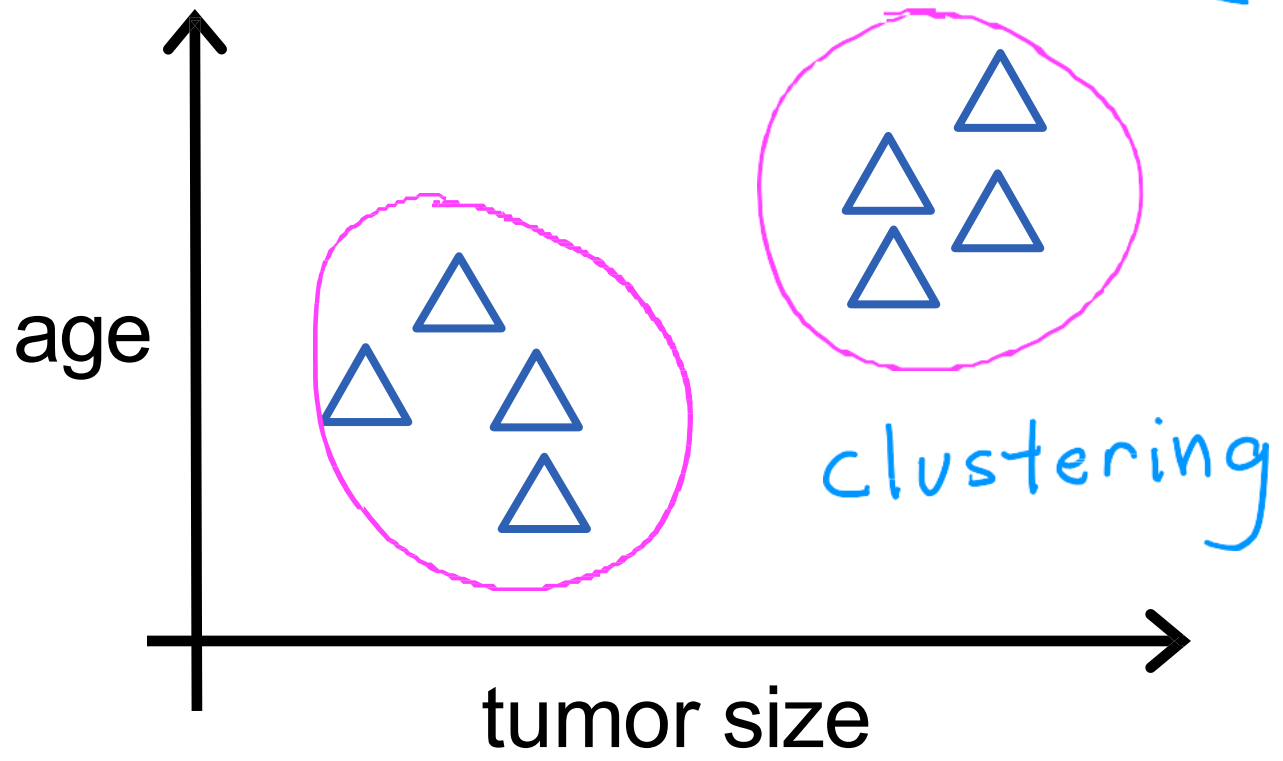
- Does not use labeled data (no Y's)
- Objective is more fuzzy:
  - Find groups of samples that behave similarly
  - Find features that behave similarly
  - Find linear combinations of features with the most variation
- Difficult to know how well you are doing
- Different from supervised learning, but can be useful as a pre-processing step for supervised learning



Supervised learning
Learn from data labeled
with the "right answers"



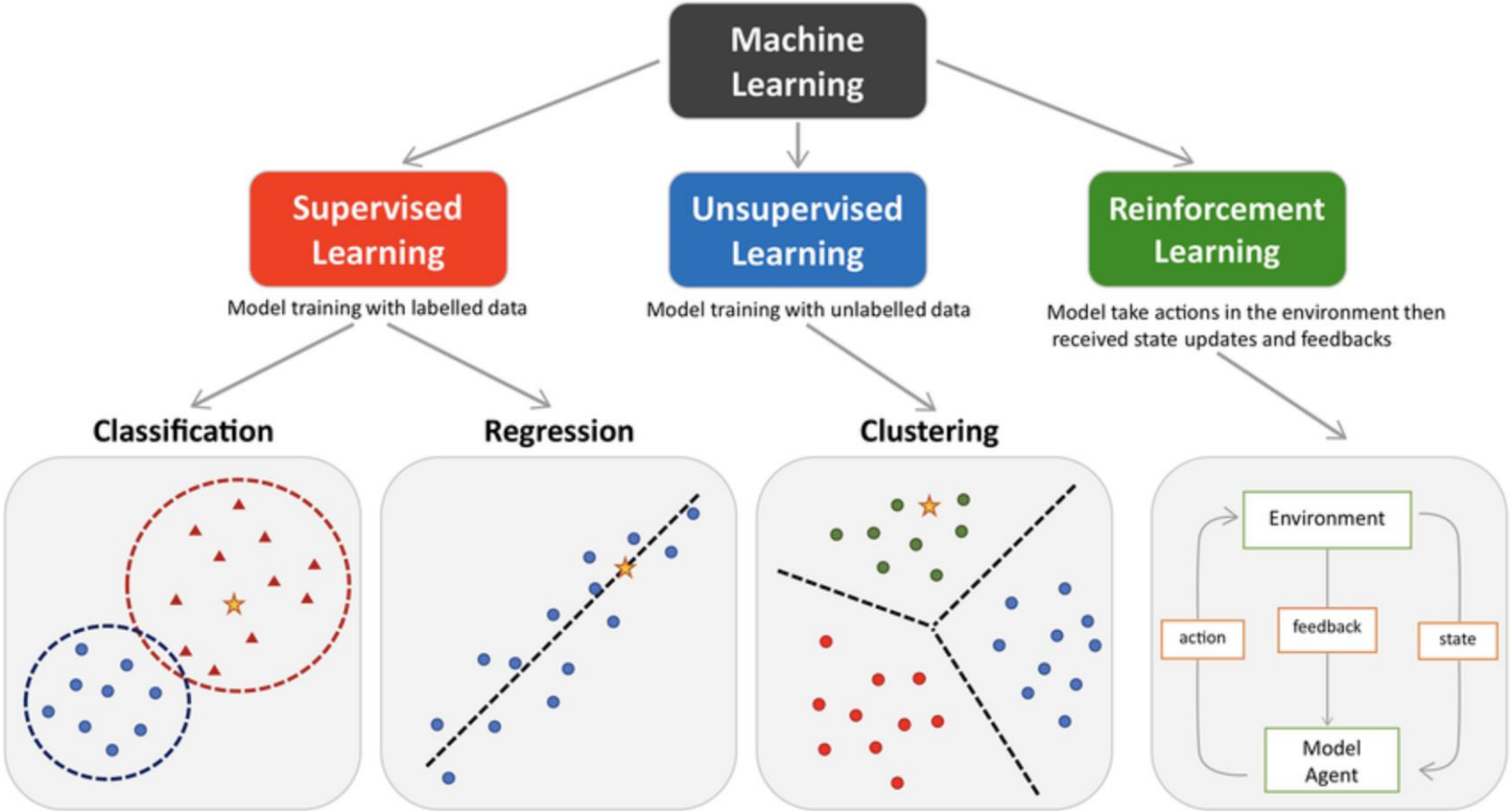
Unsupervised learning
Find something interesting
in unlabeled data.



Typical Supervised Machine Learning Flow

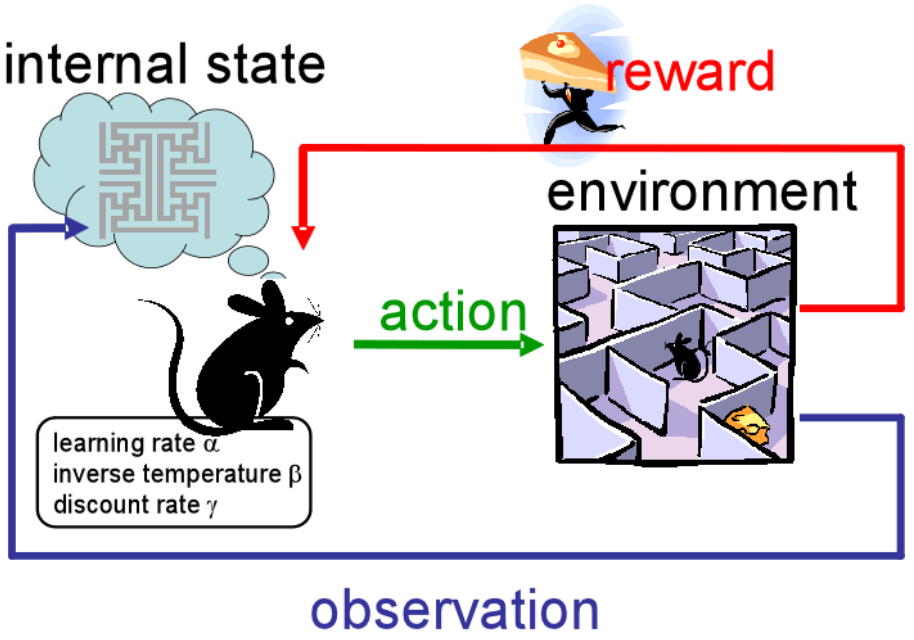
Typical Unsupervised Machine Learning Flow

Types of ML

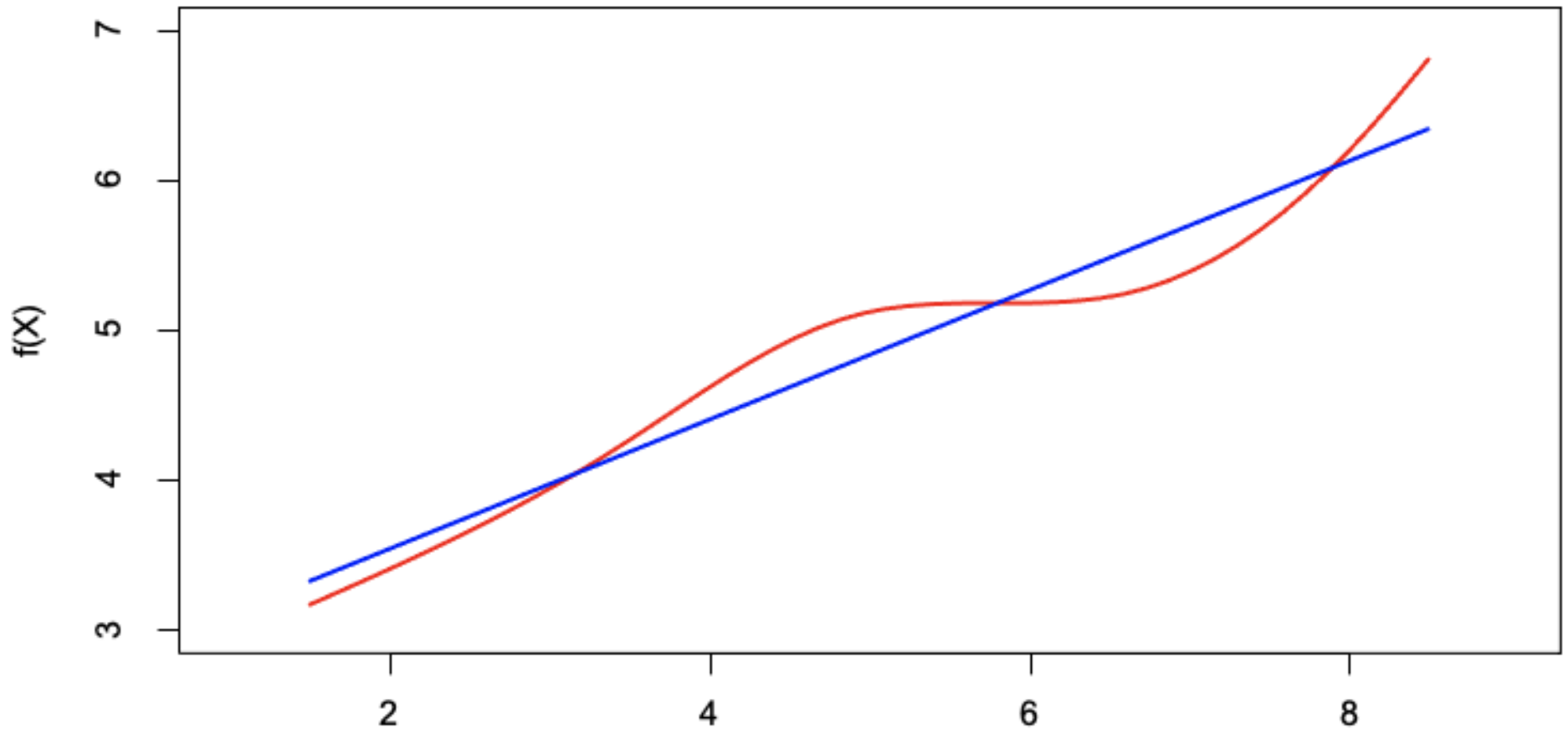

Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Model training with labelled data
Model training with unlabelled data
Model take actions in the environment then received state updates and feedbacks
Classification
Regression
Clustering
Environment
action
feedback
state
Model Agent

Reinforcement Learning

- Learns by trial and error
- Agent interacts with an environment
- Receives rewards or penalties as feedback
- Goal: maximize total reward over time
- Mainly used in:
  - Game playing (e.g., Chess, Go, video games)
  - Robotics
  - Self-driving cars
  - LLMs



Our first supervised ML algorithm



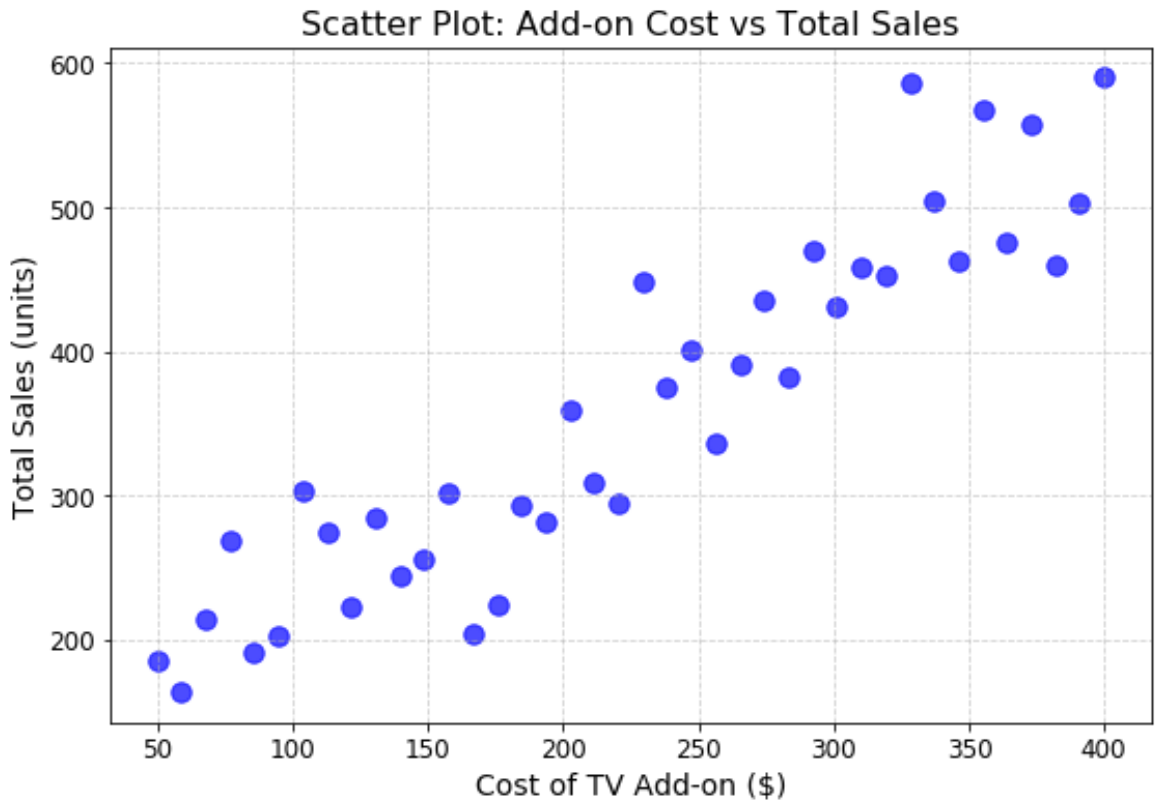
Linear Regression: Problem Statement

- Example: Predict sales from TV advertising costs
- Assumes a linear relationship between cost of the TV ad and total sales units
- Equation: $y = \beta_0 + \beta_1 x$

- X –budget, Y – total sales

Linear Regression: Data Visualization

- Scatter plot X vs. y
- Look for a linear trend



Linear Regression: Fit the Line

- Find the line of best fit
- Equation:

$$y = mx + b$$

- If we optimize m and b, we can use line to make prediction
  - How do we optimize m and b?

Linear Regression: Cost Function

- Measures “error” between predictions and actual values
- Cost function = aggregate of “error” across all data points.
- In linear regression training, we want to find the “best” line.
  - “Best” = the line that minimizes the total error between predictions and actuals

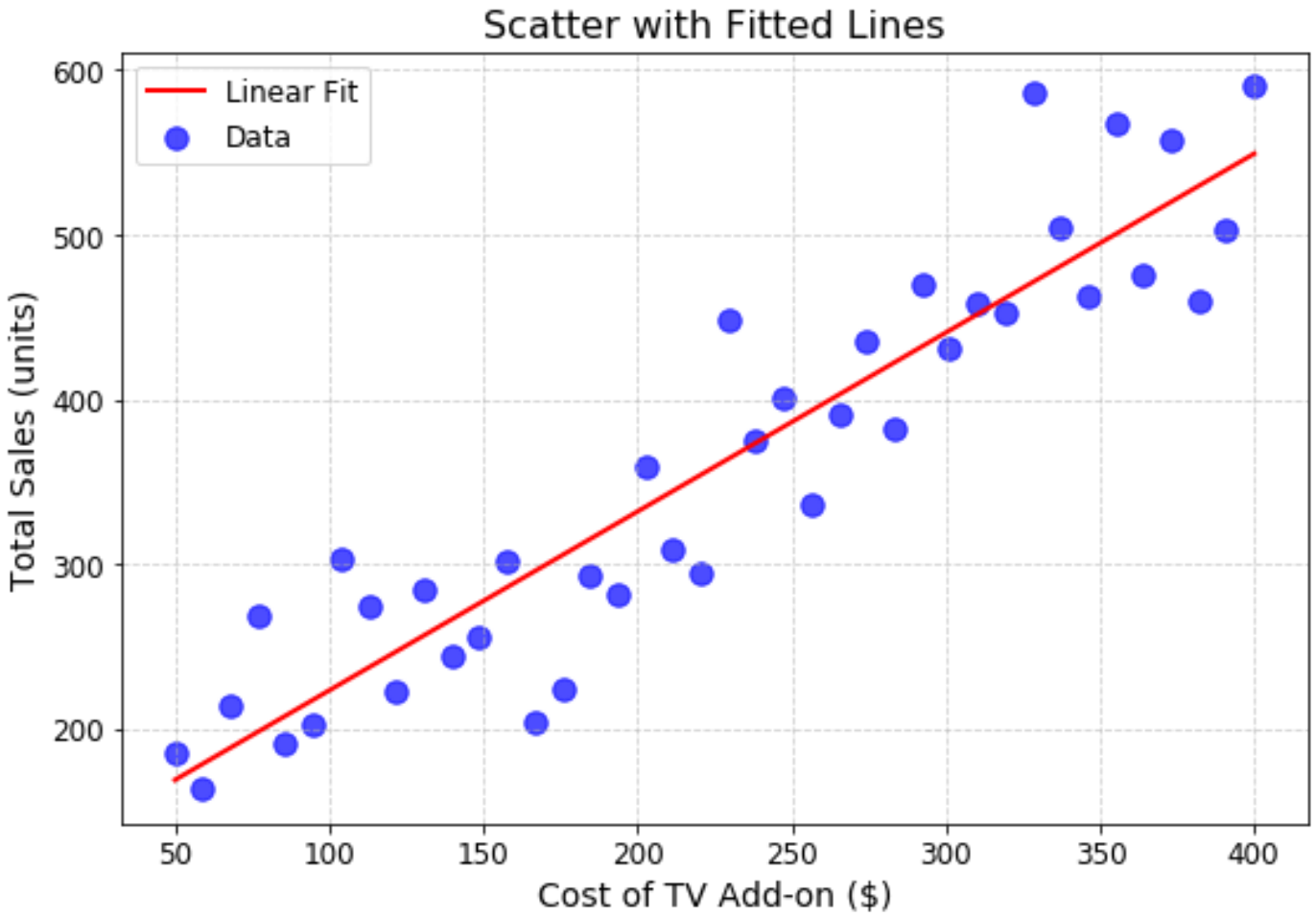
- Linear regression uses Residual Sum of Squares (RSS) to measure ”error”:

$$RSS = \sum (y_i - \hat{y}_i)^2$$

- Goal: choose coefficients that minimize RSS

Linear Regression

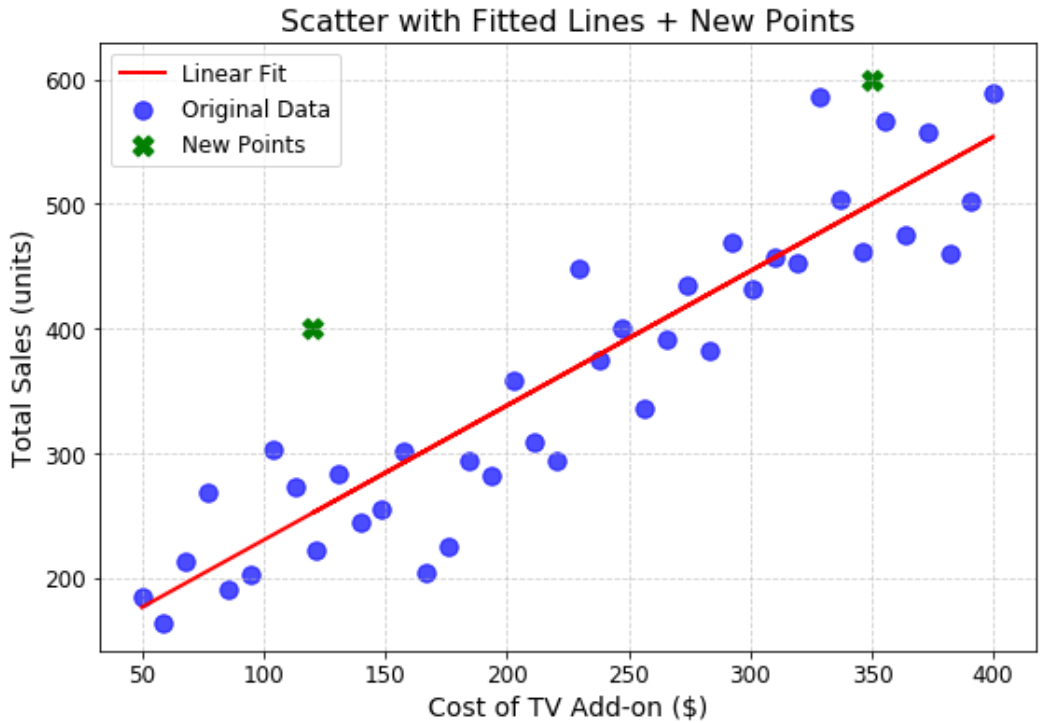
- Intuition: Line passes as close as possible to all points



Linear Regression: Evaluate the Model

- Test on new unseen data
- Compare predicted ŷ with actual y
- Metric: Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$



Linear Regression: recap

Here’s the typical workflow for supervised learning with linear regression:

- Visualize the data (scatter plot).
- Fit the line of best fit.
- Evaluate the model’s performance with metrics and test data.

We’ll later extend this process to more complex models, but the logic stays the same.

Anaconda and Github

Anaconda

- A distribution of Python & R for data science and machine learning
- Comes with pre-installed packages (NumPy, pandas, scikit-learn, etc.)
- Includes tools like Jupyter Notebook
- Helps manage environments and dependencies easily

GitHub

- A cloud platform for hosting and sharing code using Git
- Enables version control and collaboration
- Widely used for open-source projects and teamwork
- Integrates with CI/CD and project management tools

Thank you!