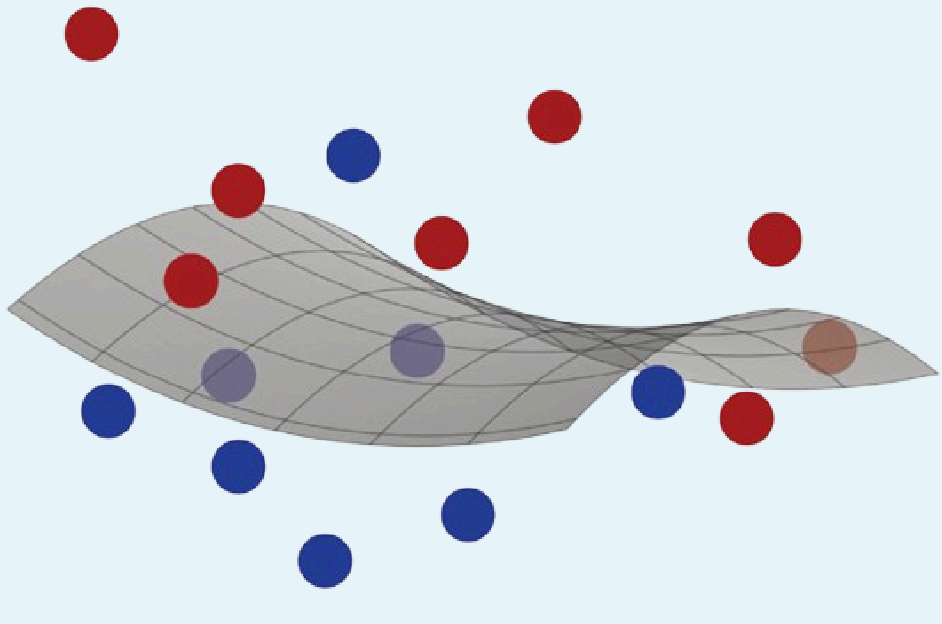
Foundations of Machine Learning

DAY - 6

Machine Learning Scenarios



Machine Learning Scenarios

Created By: **Birva Dave**
medium.com/@birva1809
github.com/Birva1809
linkedin.com/in/birva-dave
birvadave1809@gmail.com

Machine learning isn't a one-size-fits-all solution. The choice of learning approach depends on the type of data and the nature of the problem. Below are the most common learning paradigms in ML — from traditional supervised methods to more adaptive, interactive techniques.

1. Supervised Learning

- In supervised learning, the model is trained on a labeled dataset — where each example includes an input and a known output (label). The goal is to learn a mapping function that can generalize well to new, unseen data.
- Use Cases:
  - Classification (e.g., spam detection, sentiment analysis)
  - Regression (e.g., predicting house prices, temperature forecasting)
  - Ranking (e.g., ordering search engine results)
- Supervised learning is the most widely adopted ML technique, forming the backbone of many practical applications.

2. Unsupervised Learning

- Unsupervised learning uses unlabeled data — the model tries to identify hidden patterns, groupings, or structures in the data without predefined outcomes.
- Use Cases:
  - Clustering (e.g., customer segmentation, market basket analysis)
  - Dimensionality Reduction (e.g., data compression, visualization using PCA/t-SNE)
- Since there are no labels, it's more challenging to evaluate performance objectively, but it can reveal meaningful insights and relationships in raw data.

Created By: **Birva Dave**
medium.com/@birva1809
github.com/Birva1809
linkedin.com/in/birva-dave
birvadave1809@gmail.com

3. Semi-Supervised Learning

- This approach combines a small amount of labeled data with a large pool of unlabeled data. It strikes a balance between the high cost of labeling and the value of structured learning.
- Use Cases:
  - Image classification where labeling thousands of images is costly
  - Medical diagnostics using expert-labeled examples along with raw records
  - Text classification/NLP, where some documents are labeled and others are not
- The model uses the labeled data to guide learning while extracting useful structure from the unlabeled portion — often achieving better results than using labeled data alone.

4. Transductive Inference

- Here, the model is trained on labeled and unlabeled data with the test set already known. The objective is not to generalize broadly, but to make the best possible predictions only for that specific set of test examples.
- Use Cases:
  - Personalized recommendation systems (e.g., Netflix or Spotify), where users/items are fixed
- Key Characteristics:
  - Offers higher accuracy for the known test set
  - Not reusable for predicting unseen, future examples
- This targeted approach is useful when the test set is static and performance matters more than generalization.

Machine Learning Scenarios

Created By: **Birva Dave**
medium.com/@birva1809
github.com/Birva1809
linkedin.com/in/birva-dave
birvadave1809@gmail.com

5. Online Learning

- In online learning, data arrives sequentially in a stream. The model makes a prediction for each input, receives the true label immediately afterward, and updates itself based on the loss.
- Use Cases:
  - Real-time stock price prediction
  - Live recommendation engines (e.g., news feeds, shopping suggestions)
- Key Characteristics:
  - Designed for environments where data is constantly changing
  - Doesn't assume any fixed data distribution (can handle adversarial input)
  - Aims to minimize cumulative error or regret over time
- This method is ideal when learning must happen in real time, with limited memory and constant model refinement.

6. Reinforcement Learning (RL)

- In RL, an agent interacts with an environment, taking actions, receiving rewards, and observing state changes. The goal is to learn a policy that maximizes cumulative reward over time.
- Use Cases:
  - Robotics and control systems
  - Game playing (e.g., AlphaGo, Chess AI)
  - Self-driving cars
- Challenges:
  - Rewards may be delayed or sparse
  - Balancing exploration (trying new things) vs exploitation (using known strategies)
- Unlike supervised learning, RL learns through trial and error, often without a known "correct" action — making it powerful for sequential decision-making problems.

Machine Learning Scenarios

Created By: **Birva Dave**
medium.com/@birva1809
github.com/Birva1809
linkedin.com/in/birva-dave
birvadave1809@gmail.com

7. Active Learning

- Active learning lets the model decide which examples to label — by querying an oracle (usually a human expert) for the most informative data points. This is especially helpful when labeling is expensive or time-consuming.
- Use Cases:
  - Medical imaging (e.g., asking radiologists to label edge-case scans)
  - Legal document review
  - Drug discovery or bioinformatics
- Query Strategies:
  - Uncertainty sampling: Ask for labels where the model is least confident
  - Query-by-committee: Ask when multiple models disagree
  - Expected model change: Label points that would impact the model most
- By selectively labeling only the most useful data, active learning reduces annotation costs while maintaining high accuracy.

Machine Learning Scenarios

Created By: **Birva Dave**
medium.com/@birva1809
github.com/Birva1809
linkedin.com/in/birva-dave
birvadave1809@gmail.com

Type	Input Data	Goal	Common Use Cases	Key Characteristics
Supervised Learning	Labeled (input + output)	Learn mapping from input to output	Spam detection, price prediction, ranking	Most common, easy to evaluate, generalizes to unseen data
Unsupervised Learning	Unlabeled	Discover patterns or groupings	Clustering, customer segmentation, PCA	No ground truth, harder to evaluate, exploratory
Semi-Supervised	Few labeled + many unlabeled	Leverage both to improve learning	Medical diagnosis, image/text classification	Reduces labeling cost, improves performance over small labeled sets
Transductive	Labeled + unlabeled (test known)	Predict only for known test examples	Recommender systems (fixed users/items)	Optimized for a specific test set, not generalizable
Online Learning	Data in sequence (streaming)	Predict & update in real time	Stock prediction, real-time recommendations	Handles evolving data, memory-efficient, learns on-the-fly
Reinforcement Learning	Environment interaction	Maximize cumulative reward	Robotics, games (AlphaGo), autonomous vehicles	Trial-and-error learning, balances exploration & exploitation
Active Learning	Learner chooses what to label	Achieve high accuracy with fewer labels	Medical imaging, legal document review	Efficient labeling, focuses on uncertain or impactful examples