The landscape of machine learning

CSE 250B

Three learning modalities

1. **Supervised learning**
   For solving prediction problems

Three learning modalities

1. **Supervised learning**
   For solving prediction problems

2. **Unsupervised learning**
   For finding good representations

Three learning modalities

1. **Supervised learning**
   For solving prediction problems

2. **Unsupervised learning**
   For finding good representations

3. **Learning through interaction**
   E.g., reinforcement learning

Machine learning versus Algorithms

A central goal of both fields:

develop procedures that exhibit a desired input-output behavior.

- **Algorithms**: the input-output mapping can be precisely defined.
  Input: Graph G, two nodes u, v in the graph.
  Output: Shortest path from u to v in G.
- **Machine learning**: the mapping cannot easily be made precise.
  Input: Picture of an animal.
  Output: Name of the animal.

Machine learning versus Algorithms

A central goal of both fields:

develop procedures that exhibit a desired input-output behavior.

- **Algorithms**: the input-output mapping can be precisely defined.

  Input: Graph G, two nodes u, v in the graph.

  Output: Shortest path from u to v in G.

- **Machine learning**: the mapping cannot easily be made precise.

  Input: Picture of an animal.

  Output: Name of the animal.

Instead, we simply provide examples of (input,output) pairs and ask the machine to *learn* a suitable mapping itself.

Inputs and outputs

Basic terminology:

- The input space, $\mathcal{X}$.
  E.g. 32×32 RGB images of animals.
- The output space, $\mathcal{Y}$.
  E.g. Names of 100 animals.

x**:** 

y**:** "bear"

Inputs and outputs

Basic terminology:

- The input space, $\mathcal{X}$.
  E.g. 32×32 RGB images of animals.
- The output space, $\mathcal{Y}$.
  E.g. Names of 100 animals.

x**:** 

y**:** “bear”

After seeing a bunch of examples (x, y), pick a mapping

$$f : \mathcal{X} \to \mathcal{Y}$$

that accurately recovers the input-output pattern of the examples.

Inputs and outputs

Basic terminology:

- The input space, $\mathcal{X}$.
  E.g. 32×32 RGB images of animals.
- The output space, $\mathcal{Y}$.
  E.g. Names of 100 animals.

x**:** 

y**:** "bear"

After seeing a bunch of examples (x, y), pick a mapping

$$f : \mathcal{X} \to \mathcal{Y}$$

that accurately recovers the input-output pattern of the examples.

Prediction problems can be categorized by the type of **output space**: (1) discrete, (2) continuous, or (3) probability values.

Discrete output space: classification

Binary classification

E.g., Spam detection
$\mathcal{X} = \{\text{email messages}\}$
$\mathcal{Y} = \{\text{spam}, \text{not spam}\}$

Discrete output space: classification

Binary classification

E.g., Spam detection
$\mathcal{X} = \{\text{email messages}\}$
$\mathcal{Y} = \{\text{spam}, \text{not spam}\}$

Multiclass

E.g., News article classification
$\mathcal{X} = \{\text{news articles}\}$
$\mathcal{Y} = \{\text{politics}, \text{business}, \text{sports}, \ldots\}$

Discrete output space: classification

Binary classification

E.g., Spam detection
$\mathcal{X} = \{\text{email messages}\}$
$\mathcal{Y} = \{\text{spam}, \text{not spam}\}$

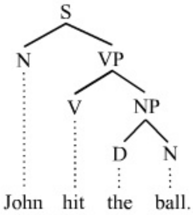
Multiclass

E.g., News article classification
$\mathcal{X} = \{\text{news articles}\}$
$\mathcal{Y} = \{\text{politics}, \text{business}, \text{sports}, \ldots\}$

Structured outputs

E.g., Parsing
$\mathcal{X} = \{\text{sentences}\}$
$\mathcal{Y} = \{\text{parse trees}\}$

$x =$ "John hit the ball"

$y =$ 

Continuous output space: regression

- Pollution level prediction
  Predict tomorrow's air quality index in my neighborhood
  $\mathcal{Y} = [0, \infty)$ $\quad$ (< 100: okay, > 200: dangerous)

- Insurance company calculations
  What is the expected life expectancy of this person?
  $\mathcal{Y} = [0, 120]$

What are suitable predictor variables ($\mathcal{X}$) in each case?

Probability estimation

$\mathcal{Y} = [0, 1]$ represents **probabilities**

Example: Credit card transactions

- $x =$ details of a transaction
- $y =$ probability this transaction is fraudulent

Probability estimation

$\mathcal{Y} = [0, 1]$ represents **probabilities**

Example: Credit card transactions

- x = details of a transaction
- y = probability this transaction is fraudulent

Why not just treat this as a binary classification problem?

Three learning modalities

1. **Supervised learning**

   Methods:
   nearest neighbor, generative models for prediction, linear regression, logistic regression, perceptron, support vector machines, kernel methods, decision trees, boosting, random forests, neural nets

   Underlying math:
   linear algebra, optimization, probability

   Formal models:
   statistical learning framework, online learning

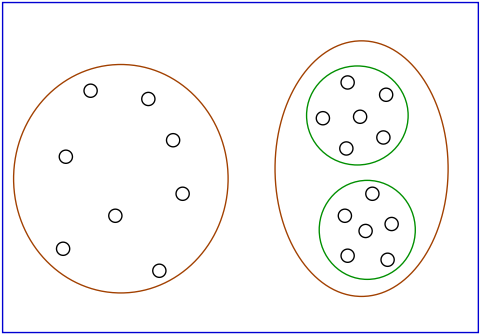
2. **Unsupervised learning**
3. **Learning through interaction**

Unsupervised learning

Find **structure** in data: underlying **degrees of freedom**.

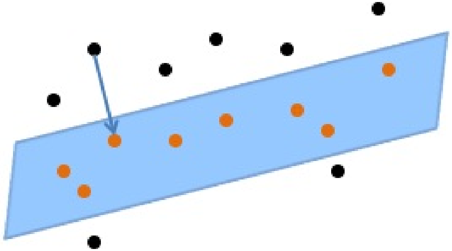
Unsupervised learning

Find **structure** in data: underlying **degrees of freedom**.



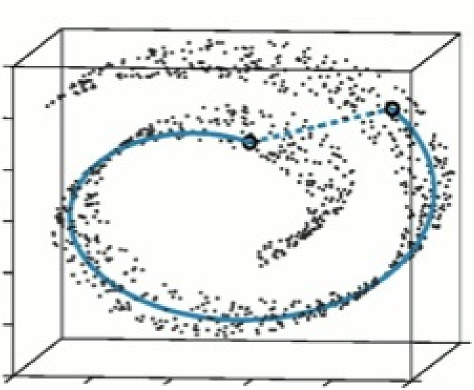
Unsupervised learning

Find **structure** in data: underlying **degrees of freedom**.



Unsupervised learning

Find **structure** in data: underlying **degrees of freedom**.



Three learning modalities

1. **Supervised learning**

2. **Unsupervised learning**

   Types of structure:
   clusters; low-dimensional subspaces; manifolds; dictionaries; independent components; topics

   Algorithmic foundations:
   local search; linear algebra

3. **Learning through interaction**