### The learning problem - Outline

- Example of machine learning
- Components of Learning
- A simple model
- Types of learning
- Puzzle

D

**Example:** Predicting how a viewer will rate a movie

**10%** improvement = 1 million dollar prize

The essence of machine learning:

- A pattern exists.
- We cannot pin it down mathematically.
- We have data on it.

D



The learning approach



### Components of learning

### Metaphor: Credit approval

Applicant information:

age	23 years
gender	male
annual salary	\$30,000
years in residence	1 year
years in job	1 year
current debt	\$15,000
• • •	• • •

#### Approve credit?

### Components of learning

### Formalization:

- Input: **x** (customer application)
- Output: *y* (good/bad customer?)
- Target function:  $f : \mathcal{X} \to \mathcal{Y}$  (ideal credit approval formula)

• Data: 
$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)$$
 (historical records)  
 $\downarrow \quad \downarrow \quad \downarrow$ 

• Hypothesis:  $g: \mathcal{X} \to \mathcal{Y}$  (formula to be used)



### Solution components

The 2 solution components of the learning problem:

• The Hypothesis Set

 $\mathcal{H} = \{h\} \qquad g \in \mathcal{H}$ 

• The Learning Algorithm

Together, they are referred to as the *learning model*.



A simple hypothesis set - the 'perceptron'

For input  $\mathbf{x} = (x_1, \cdots, x_d)$  'attributes of a customer'

Approve credit if 
$$\sum_{i=1}^d w_i x_i > \text{threshold},$$
  
Deny credit if  $\sum_{i=1}^d w_i x_i < \text{threshold}.$ 

This linear formula  $h \in \mathcal{H}$  can be written as

$$h(\mathbf{x}) = \operatorname{sign}\left(\left(\sum_{i=1}^{d} w_{i} x_{i}\right) - \operatorname{threshold}\right)$$

$$h(\mathbf{x}) = \operatorname{sign}\left(\left(\sum_{i=1}^{d} \boldsymbol{w}_{i} x_{i}\right) + \boldsymbol{w}_{0}\right)$$

Introduce an artificial coordinate  $x_0 = 1$ :

$$h(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=0}^{d} \boldsymbol{w_i} \; x_i\right)$$

In vector form, the perceptron implements

 $h(\mathbf{x}) = \operatorname{sign}(\mathbf{w}^{\mathsf{T}}\mathbf{x})$ 





'linearly separable' data

A simple learning algorithm - PLA

The perceptron implements

$$h(\mathbf{x}) = \operatorname{sign}(\mathbf{w}^{\mathsf{T}}\mathbf{x})$$

Given the training set:

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\cdots,(\mathbf{x}_N,y_N)$$

pick a misclassified point:

$$sign(\mathbf{w}^{\mathsf{T}}\mathbf{x}_n) \neq y_n$$

and update the weight vector:

$$\mathbf{w} \leftarrow \mathbf{w} + y_n \mathbf{x}_n$$





### Iterations of PLA

• One iteration of the PLA:

 $\mathbf{w} \leftarrow \mathbf{w} + y\mathbf{x}$ 

where  $(\mathbf{x}, y)$  is a misclassified training point.

• At iteration  $t = 1, 2, 3, \cdots$ , pick a misclassified point from

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)$$

and run a PLA iteration on it.



• That's it!

### The learning problem - Outline

- Example of machine learning
- Components of learning
- A simple model
- Types of learning
- Puzzle

D

### Supervised vs Unsupervised Learning

### What is Supervised Machine Learning?

In Supervised learning, you train the machine using data which is well **"labeled**." It means some data is already tagged with the correct answer. It can be compared to learning which takes place in the presence of a supervisor or a teacher.

A supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data. Successfully building, scaling, and deploying accurate supervised machine learning Data science model takes time and technical expertise from a team of highly skilled data scientists. Moreover, Data scientist must rebuild models to make sure the insights given remains true until its data changes.

#### Supervised vs Unsupervised Learning

### What is Unsupervised Learning?

Unsupervised learning is a machine learning technique, where you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabelled data.

Unsupervised learning algorithms allow you to perform more complex processing tasks compared to supervised learning. Although, unsupervised learning can be more unpredictable compared with other natural learning deep learning and reinforcement learning methods.

Supervised vs Unsupervised Learning

### Why Supervised Learning?

Supervised learning allows you to collect data or produce a data output from the previous experience.

Helps you to optimize performance criteria using experience

Supervised machine learning helps you to solve various types of real-world computation problems.

#### Why Unsupervised Learning?

Unsupervised machine learning finds all kind of unknown patterns in data.

Unsupervised methods help you to find features which can be useful for categorization.

It is taken place in real time, so all the input data to be analyzed and labeled in the presence of learners.

It is easier to get unlabeled data from a computer than labeled data, which needs manual intervention.

Supervised vs Unsupervised Learning

How Supervised Learning works

For example, you want to train a machine to help you predict how long it will take you to drive home from your workplace. Here, you start by creating a set of labeled data. This data includes

Weather conditions Time of the day Holidays

All these details are your inputs. The output is the amount of time it took to drive back home on that specific day.

Supervised vs Unsupervised Learning

How Supervised Learning works



Let's see now how you can develop a supervised learning model of this example which help the user to determine the commute time. The first thing you requires to create is a training data set. This training set will contain the total commute time and corresponding factors like weather, time, etc. Based on this training set, your machine might see there's a direct relationship between the amount of rain and time you will take to get home.

So, it ascertains that the more it rains, the longer you will be driving to get back to your home. It might also see the connection between the time you leave work and the time you'll be on the road.

The closer you're to 6 p.m. the longer time it takes for you to get home. Your machine may find some of the relationships with your labeled data.

### Supervised vs Unsupervised Learning

#### How Unsupervised Learning works Let's, take the case of a baby and her family dog.



She knows and identifies this dog. A few weeks later a family friend brings along a dog and tries to play with the baby.

Supervised vs Unsupervised Learning

How Unsupervised Learning works



Baby has not seen this dog earlier. But it recognizes many features (2 ears, eyes, walking on 4 legs) are like her pet dog. She identifies a new animal like a dog. This is unsupervised learning, where you are not taught but you learn from the data (in this case data about a dog.) Had this been supervised learning, the family friend would have told the baby that it's a dog.

### Supervised vs Unsupervised Learning

### **Types of Supervised Machine Learning Techniques**



#### **Regression:**

Regression technique predicts a single output value using training data.

Example: You can use regression to predict the house price from training data. The input variables will be locality, size of a house, etc.

### **Classification:**

Classification means to group the output inside a class. If the algorithm tries to label input into two distinct classes, it is called binary classification. Selecting between more than two classes is referred to as multiclass classification.

Supervised vs Unsupervised Learning

**Types of Unsupervised Machine Learning Techniques** 

### Clustering



sample

Cluster/group

Clustering is an important concept when it comes to unsupervised learning. It mainly deals with finding a structure or pattern in a collection of uncategorized data. Clustering algorithms will process your data and find natural clusters(groups) if they exist in the data. You can also modify how many clusters your algorithms should identify. It allows you to adjust the granularity of these groups.

Supervised vs Unsupervised Learning

**Types of Unsupervised Machine Learning Techniques** 

#### Association

Association rules allow you to establish associations amongst data objects inside large databases. This unsupervised technique is about discovering exciting relationships between variables in large databases. For example, people that buy a new home most likely to buy new furniture.

Other Examples:

- A subgroup of cancer patients grouped by their gene expression measurements
- Groups of shopper based on their browsing and purchasing histories
- Movie group by the rating given by movies viewers

	Supervised VS	Unsupervised
Parameters	Supervised machine learning technique	Unsupervised machine learning technique
Process	In a supervised learning model, input and output variables will be given.	In unsupervised learning model, only input data will be given
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data which is not labeled
Algorithms Used	Support vector machine, Neural network, Linear and logistics regression, random forest, and Classification trees.	Unsupervised algorithms can be divided into different categories: like Cluster algorithms, K-means, Hierarchical clustering, etc.
Computational Complexity	Supervised learning is a simpler method.	Unsupervised learning is computationally complex
Use of Data	Supervised learning model uses training data to learn a link between the input and the outputs.	Unsupervised learning does not use output data.
Accuracy of Results	Highly accurate and trustworthy method.	Less accurate and trustworthy method.
Real Time Learning	Learning method takes place offline.	Learning method takes place in real time.
Number of Classes	Number of classes is known.	Number of classes is not known.
Main Drawback	Classifying big data can be a real challenge in Supervised Learning.	You cannot get precise information regarding data sorting, and the output as data used in unsupervised learning is labeled and not known.

Supervised vs Unsupervised Learning

### Summary

- In Supervised learning, you train the machine using data which is well "labeled."
- Unsupervised learning is a machine learning technique, where you do not need to supervise the model.
- Supervised learning allows you to collect data or produce a data output from the previous experience.
- Unsupervised machine learning helps you to finds all kind of unknown patterns in data.
- For example, you will able to determine the time taken to reach back come base on weather condition, Times of the day and holiday.
- For example, Baby can identify other dogs based on past supervised learning.
- Regression and Classification are two types of supervised machine learning techniques.
- Clustering and Association are two types of Unsupervised learning.
- In a supervised learning model, input and output variables will be given while with unsupervised learning model, only input data will be given



### Dermatologist-level classification of skin cancer

An artificial intelligence trained to classify images of skin lesions as benign lesions or malignant skin cancers achieves the accuracy of board-certified dermatologists.

In this work, we pretrain a deep neural network at general object recognition, then finetune it on a dataset of ~130,000 skin lesion images comprised of over 2000 diseases.

FULL NATURE ARTICLE >

OPEN-ACCESS PDF >