Lecture 1

Machine Learning Basics

Dong Kook Kim

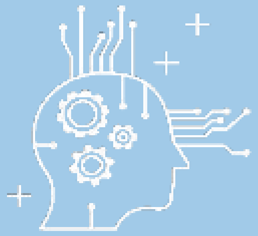
Basic ML Concepts

- What is DL, ML and AI
- What is Learning ?
  - Supervised
  - Unsupervised
- What is Regression?
- What is Classification?
- 5 Steps for Learning

What is DL, ML, and AI

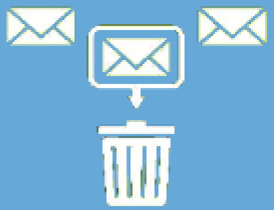
Artificial Intelligence

Any technique that enables computers to mimic human behavior



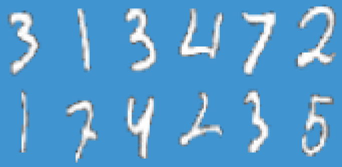
Machine Learning

Ability to learn without explicitly being programmed

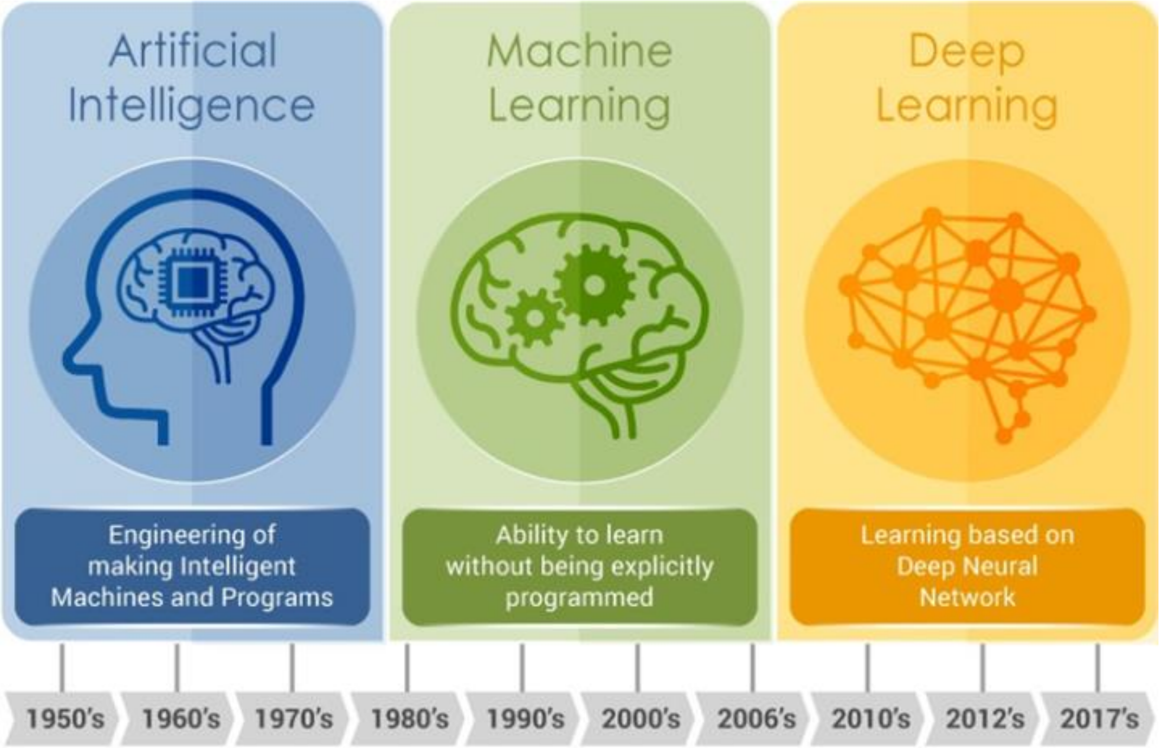


Deep Learning

Extract patterns from data using neural networks

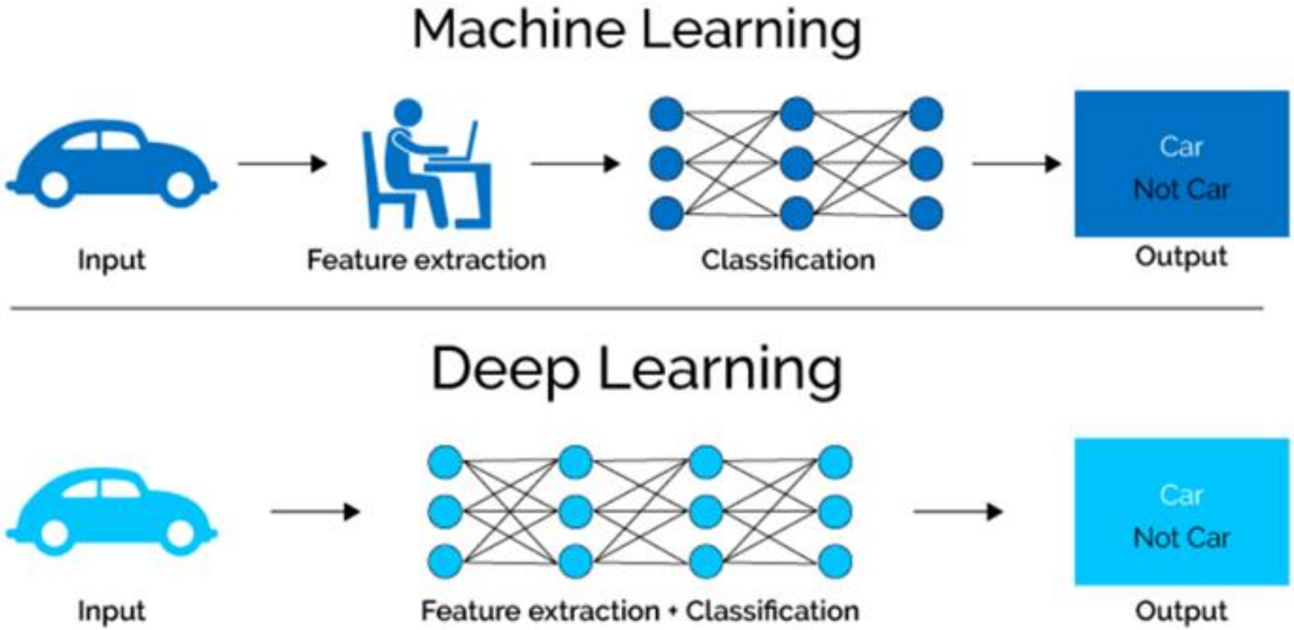


What is DL, ML, and AI



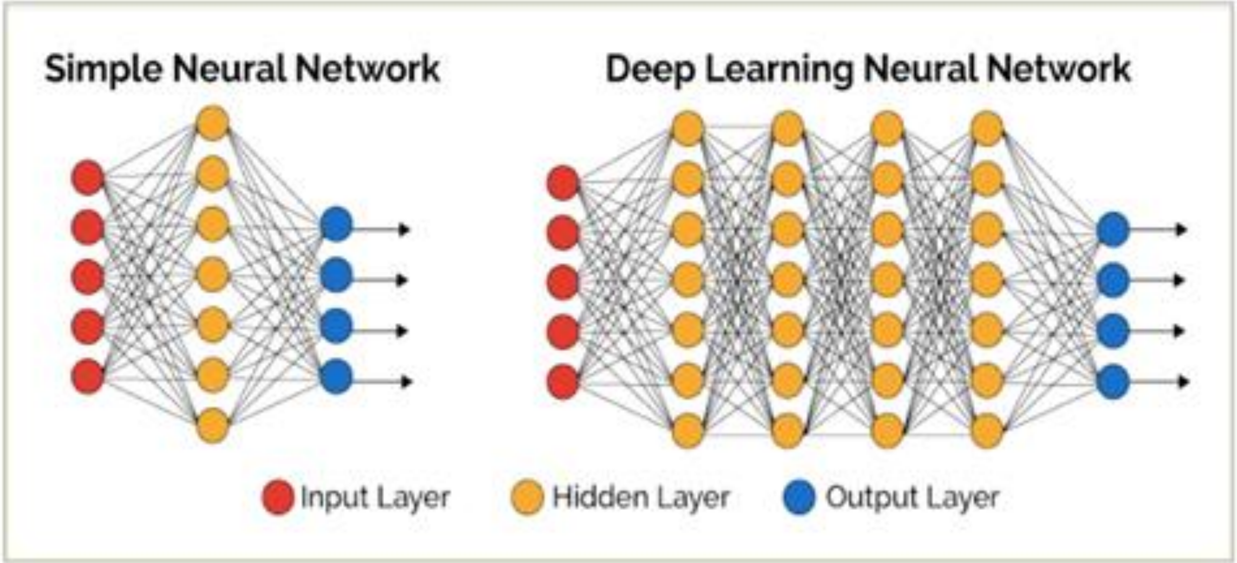
https://towardsdatascience.com/notes-on-artificial-intelligence-ai-machine-learning-ml-and-deep-learning-dl-for-56e51a2071c2/

ML vs DL



https://towardsdatascience.com/notes-on-artificial-intelligence-ai-machine-learning-ml-and-deep-learning-dl-for-56e51a2071c2/

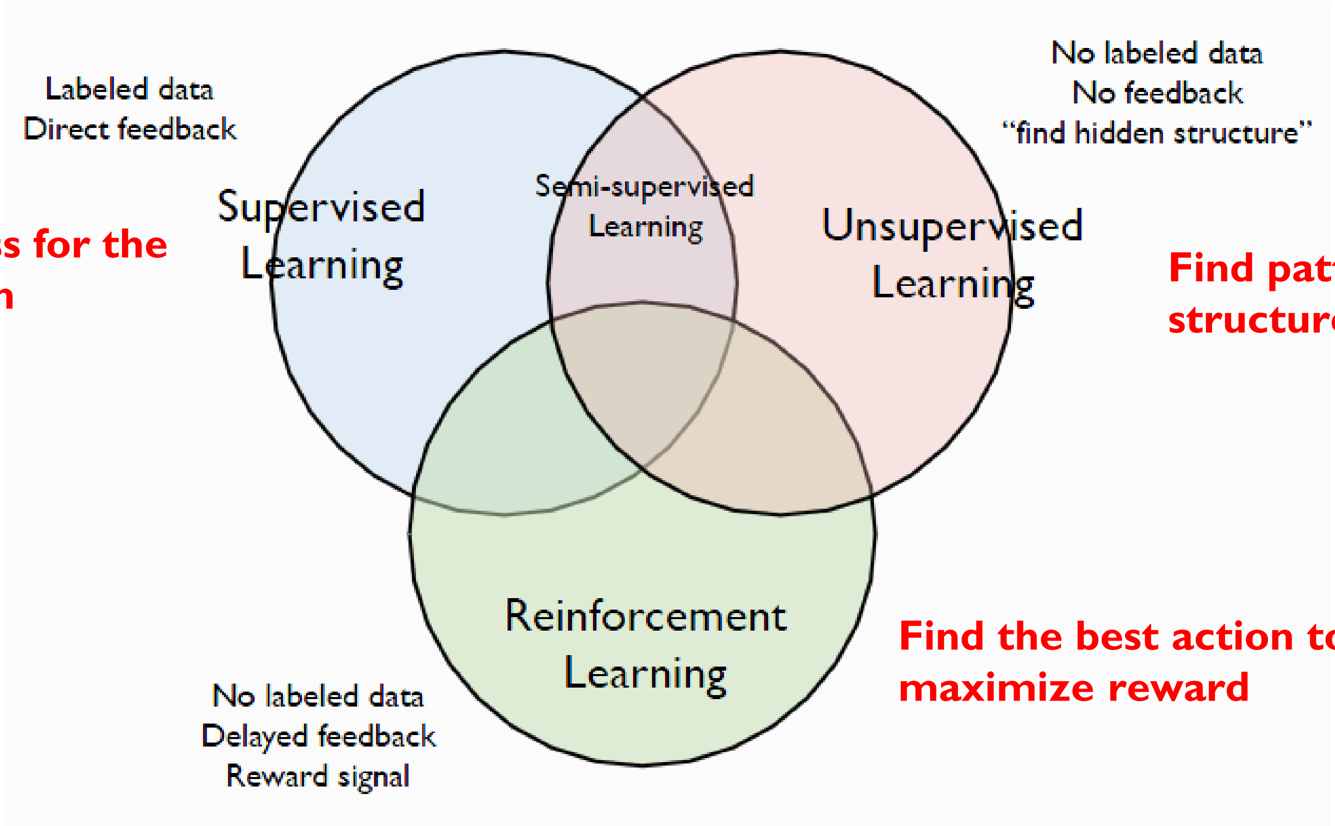
What is the “Deep” in DL ?



https://towardsdatascience.com/notes-on-artificial-intelligence-ai-machine-learning-ml-and-deep-learning-dl-for-56e51a2071c2/

Types of ML

Find a class for the input given

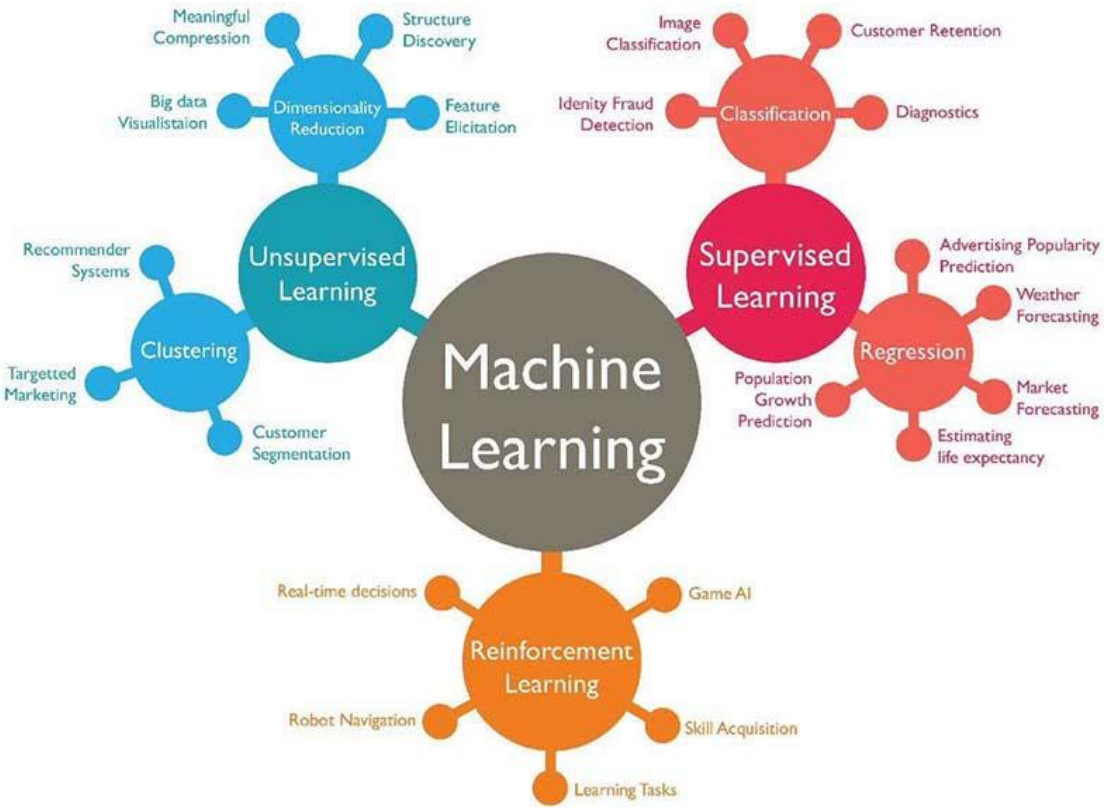


Find patterns or structure in data

Find the best action to maximize reward

https://www.techleer.com/articles/203-machine-learning-algorithm-backbone-of-emerging-technologies/

Types of ML



https://www.techleer.com/articles/203-machine-learning-algorithm-backbone-of-emerging-technologies/

Learning Algorithms

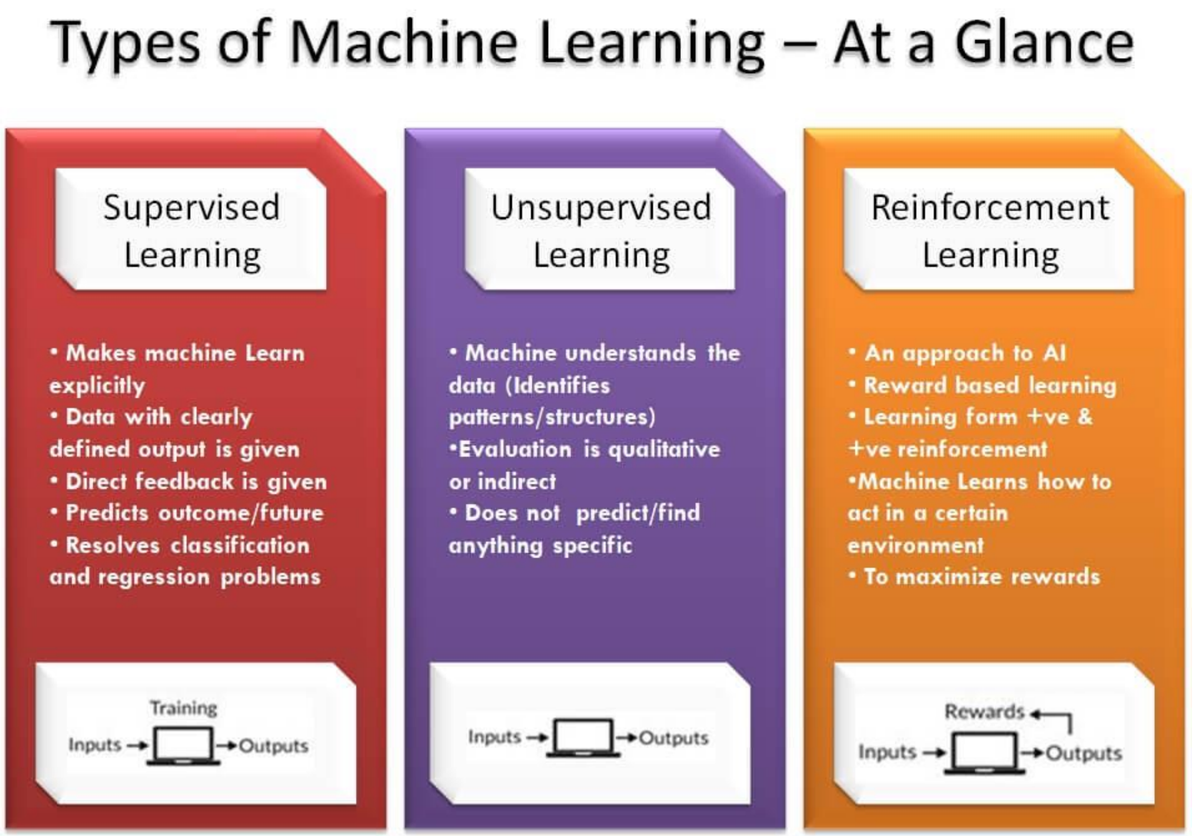
- Supervised learning
  - learning with labeled data
  - given input x and target y
  - $y = f(x)$: predict target y corresponding to input x

- Unsupervised learning
  - learning with un-labeled data
  - given only input x
  - $f(x)$: Estimate the distribution of input x
  - $z = f(x)$: Find the latent variable z given input x

Learning Algorithms

- Reinforcement Learning
  - learning with un-labelled
  - given input x

  - $y = f(x)$: Predict action y based on input x to maximize a future reward z
    z

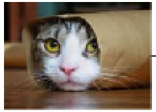


https://www.newtechdojo.com/list-machine-learning-algorithms/

Supervised learning

An example training set for four visual categories

- Learn a function **f** to map x → y

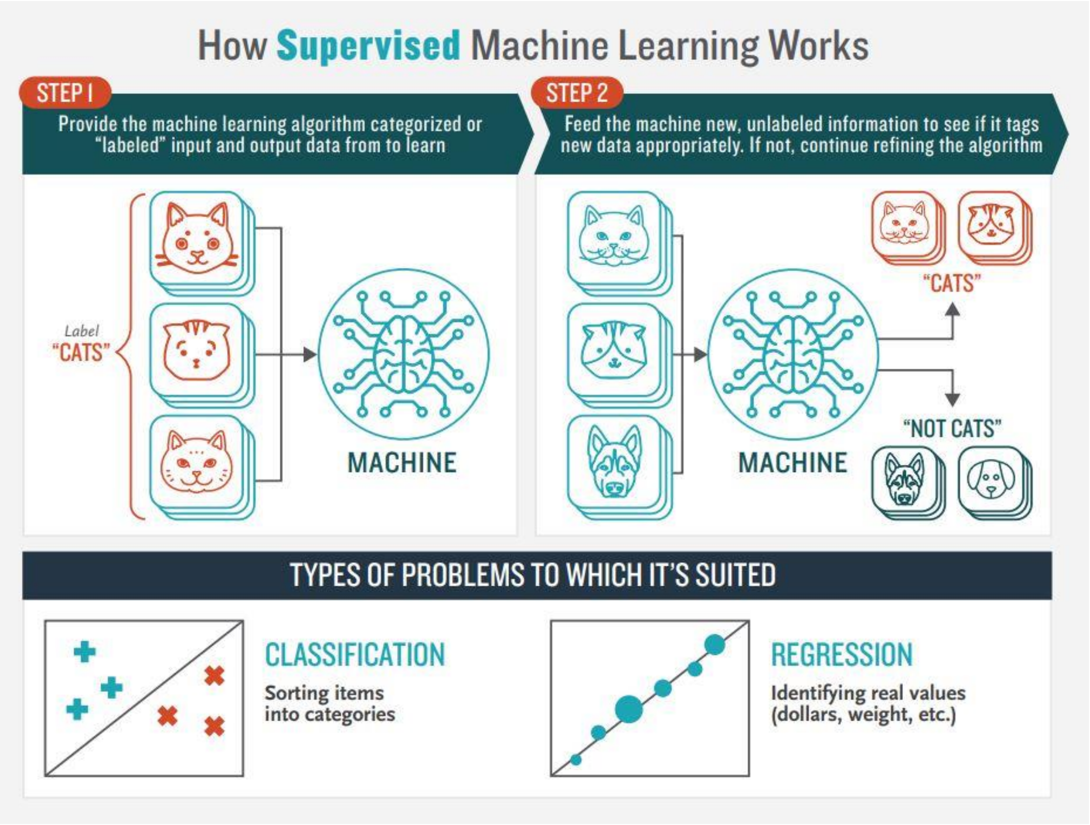


$\rightarrow \boxed{f} \rightarrow$

Cat : 0.98
Cake : 0.02
Dog : 0.00

- y = f(x)

Supervised learning



https://www.newtechdojo.com/list-machine-learning-algorithms/

Supervised learning

- Most common problem type in ML

  - Image labeling : learning from tagged images
  - Email spam filter : learning from labeled (spam or ham email)
  - Predicting exam score : learning from previous exam score and time spent

Types of supervised learning

- Predicting final exam score based on time spent
  - Regression (target : real)
- Pass / Non-pass based on time spent
  - Binary classification (target : binary)
- Letter grade (A, B, C, E and F) based on time spent
  - Multi-label classification (target : integer)

Predicting final exam score based on time spent

- input - target

- regression

x (hours)	y (score)
10	90
9	80
3	50
2	30

Pass / Non-pass based on time spent

- binary classification

- input — target

x (hours)	y (pass/fail)
10	P
9	P
3	F
2	F

Letter grade (A, B, C, E and F) based on time spent

- multi-label classification

- input   - target

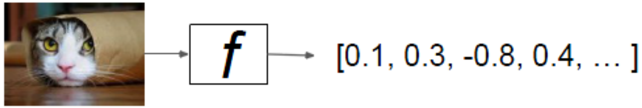
x (hours)	y (grade)
10	A
9	B
3	D
2	F

A : 1
B : 2
C : 3
D : 4
E : 5
F : 6

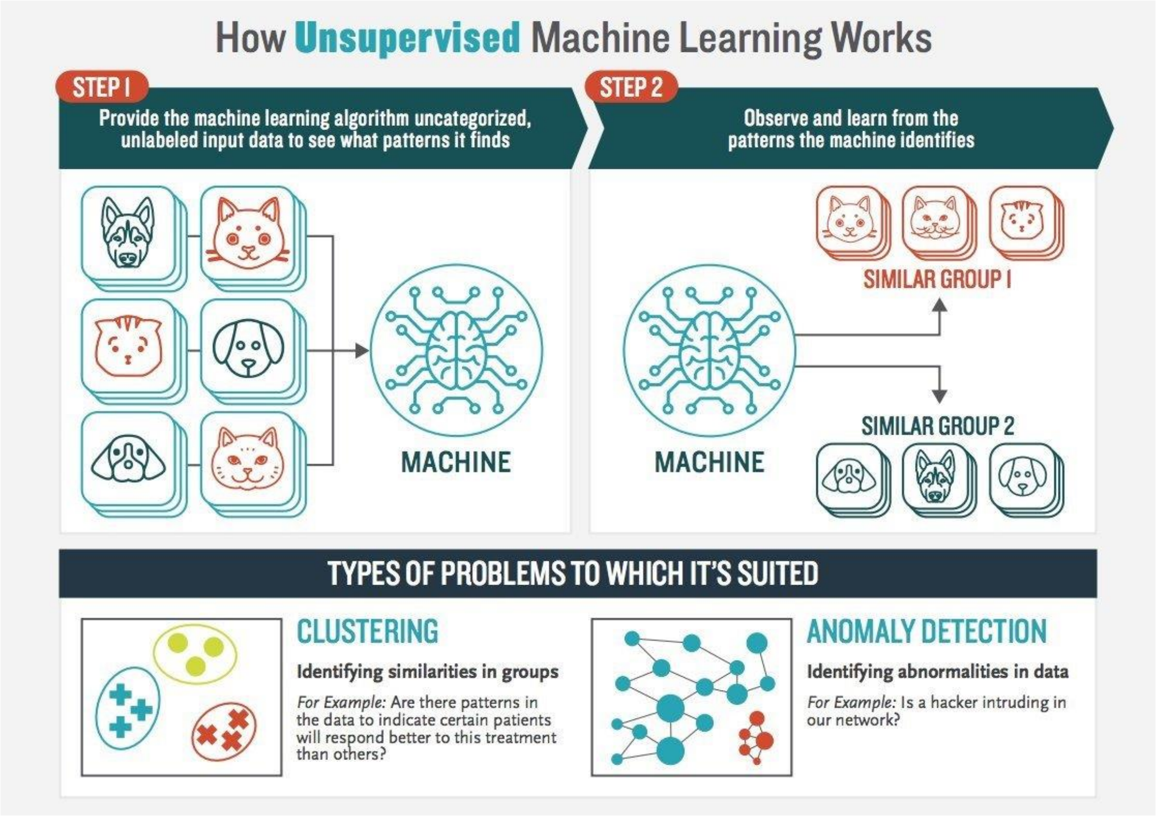
Unsupervised learning

An example training set (no labels)

- Learn a function **f** to find some underlying hidden structure of the data x



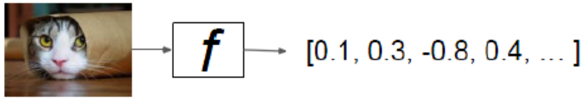
Unsupervised learning



https://www.newtechdojo.com/list-machine-learning-algorithms/

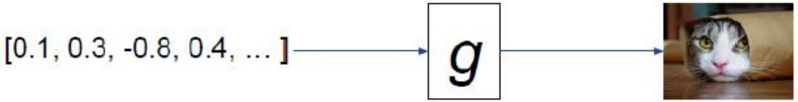
Types of Unsupervised learning

- Feature learning : given x, find a new feature z



  - z = f(x)

- Generative modeling : given z, generate a new x



  - x = g(z)

5 Steps for Learning

Step 1. Training examples, $(x^{(i)}, y^{(i)})$ or $x^{(i)}$ only
$x^{(i)}$: input feature (vectors), $y^{(i)}$: target

Step 2. A model, a function that represents the relationship between x and y
$y = f(x)$ or, a function that models x, $f(x)$, with parameters θ

Step 3. A loss or a cost or an objective function $C(\theta)$, which tells us how well our model approximates the training examples

Step 4. Optimization, a way of finding the parameters of our model that minimizes the loss function

Step 5. Testing, performance evaluation using test examples

Inference and Decision

- 2 Classification Stage

1. Inference stage in which we use training data to learn a model $y=f(x)$
   - Step 1. ~ Step 4.
   - Learning, Training Phase

2. Decision stage in which we use these model to make optimal class assignments
   - Step 5.
   - Testing Phase

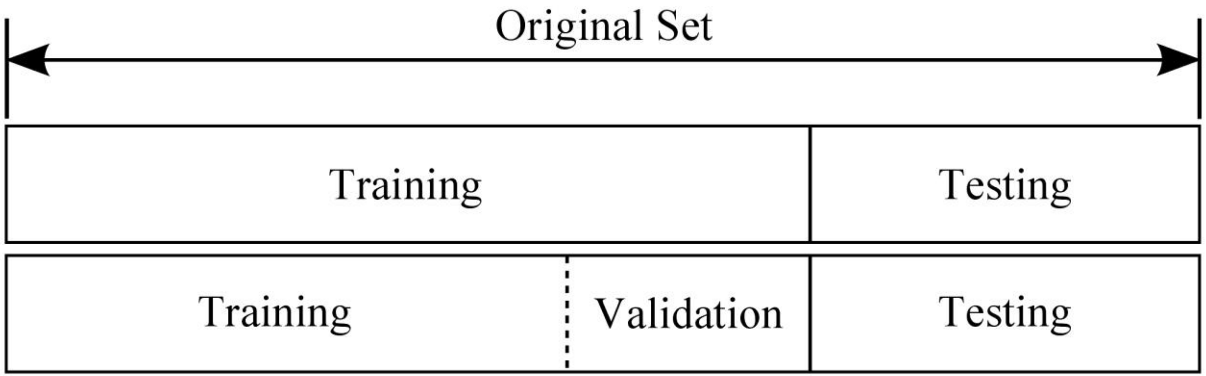
Step 1. Data

- Application dependent

- Examples
  - MNIST (handwritten digits recognition)
  - IMAGENET (object recognition)
  - WSJ (speech recognition)
  - Text etc.

Training, Validation, and Test DataSets

- Training : The sample of data used to fit the model.
- Validation : The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.
- Test : The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.



https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7

Full-batch, mini-batch

- Full batch : All data
- Mini-batch : Subsets of data

Data Representation

	Scalar	Vector	Matrix	3 Tensor	4 Tensor
Binary	0, 1	Gray image - a reshaped image (28*28=784-dim)	Gray image - (28, 28) - (N, 784)	Batch gray image - (N, 28, 28)	
Integer	1,2,.., 10	Text - A word - One-hot encoded	Text - A word seq.	Text - Batch word seq.	
Real	3.14	Speech - A frame	Speech - Batch frame	Color image - RGB (28, 28, 3)	Batch color image - Batch RGB - (N, 28, 28, 3)

N : batch size
Gray image : MNIST, Color image : CIFAR-10

MNIST Data

10 classes, 60000 training images, 10000 test images

```
train-images-idx3-ubyte.gz:   training set images (9912422 bytes)
train-labels-idx1-ubyte.gz:   training set labels (28881 bytes)
t10k-images-idx3-ubyte.gz:    test set images (1648877 bytes)
t10k-labels-idx1-ubyte.gz:    test set labels (4542 bytes)
```

http://yann.lecun.com/exdb/mnist/

Step 2. Models

- Supervised learning

  1. Linear models
     - Linear regression
     - Linear classification (logistic/multinomial)

  2. Nonlinear models (Neural Networks)
     - Feed-forward neural network (FFNN)
     - Convolutional neural network (CNN)
     - Recurrent neural network (RNN)

Step 2. Models

- Unsupervised learning

  1. Feature Learning
     - Autoencoder (AE)

  2. Generative Modeling
     - Variational autoencoder (VAE)
     - Generative adversarial network (GAN)

Step 2. Models

- Deterministic/Probabilistic models

Models		Deterministic	Probabilistic
Supervised	Regression	$\mathrm{y} = f_\theta(\mathrm{x})$	$p_\theta(\mathrm{y}\|\mathrm{x})$
	Binary	$y = f_\theta(\mathrm{x})$	$p_\theta(y\|\mathrm{x})$
	Multi-class (K-class)	$y_1 = f_{1,\theta}(\mathrm{x})$: $y_K = f_{K,\theta}(\mathrm{x})$	$p_\theta(y = 1\|\mathrm{x})$: $p_\theta(y = K\|\mathrm{x})$
Unsupervised		$f_\theta(\mathrm{x})$	$p_\theta(\mathrm{x})$ or $p_\theta(\mathrm{x}, \mathrm{z})$

$f_\theta(\mathrm{x})$: functions, $p_\theta(\mathrm{y}|\mathrm{x})$: conditional pdf/pmf,
$p_\theta(\mathrm{x})$: pdf/pmf, $p_\theta(\mathrm{x}, \mathrm{z})$: joint pdf/pmf

x : input, y : target, θ : model parameters

Step 3. Loss functions

- Cost function $C(\theta)$: tells us how well our model fits the training data
  - function of parameters θ

1. Regression
   - sum of square error function, or
   - MSE (mean square error)

2. Classification
   - cross-entropy function (negative log-likelihood function)
   - Binary or multi-class

Step 4. Optimization

- Optimization : determines how the network will be updated the parameters based on the loss function

- Gradient descent (GD) optimization
- Optimizers
  1. SGD
  2. AdaGrad
  3. RMSProp
  4. Adam
  5. etc.

Step 5. Testing

- Decision rules

Models		Deterministic	Probabilistic
Supervised	Regression	$\hat{\mathrm{y}} = f_\theta(\mathrm{x})$	$\hat{y} = E[\mathrm{y}\vert\mathrm{x}]$
	Binary	if $y = f_\theta(\mathrm{x}) > T : 1$ else 0	if $p_\theta(y = 1\vert\mathrm{x}) > p_\theta(y = 1\vert\mathrm{x})$: 1 else 0
	Multi-class (K-class)	*Decide class k* $k = max_j\ f_{j,\theta}(\mathrm{x})$	*Decide class k* $k = max_j\ p_\theta(y = j\vert\mathrm{x})$
Unsupervised		$z = f_\theta(\mathrm{x})$	$z = p_\theta(\mathrm{x})$ or $x = p_\theta(\mathrm{z})$
x : new input, $\hat{y}$: predicted target, θ : model parameter			

Step 5. Performance Metrics

- Metrics : function that is used to judge the performance of the model

1. Metrics for regression : MSE, or MAE

2. Metrics for classification : Accuracy (%) or Error rate (%)

3. Metrics for unsupervised : log-likelihood etc.