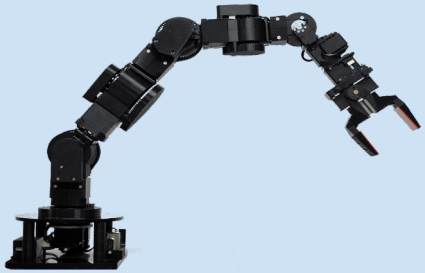
Introduction to deep learning

1.1 AI, ML & DL

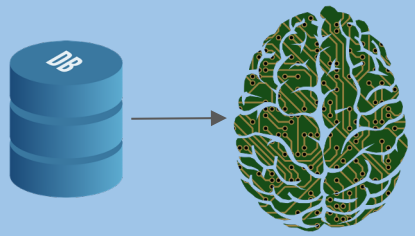
Artificial Intelligence



a very broad field

1950s

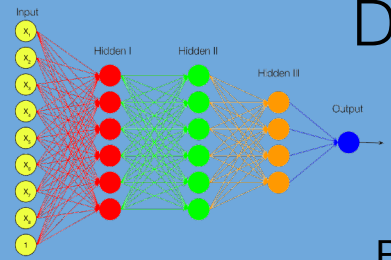
Machine Learning



Fundamentals of
“learning from data”

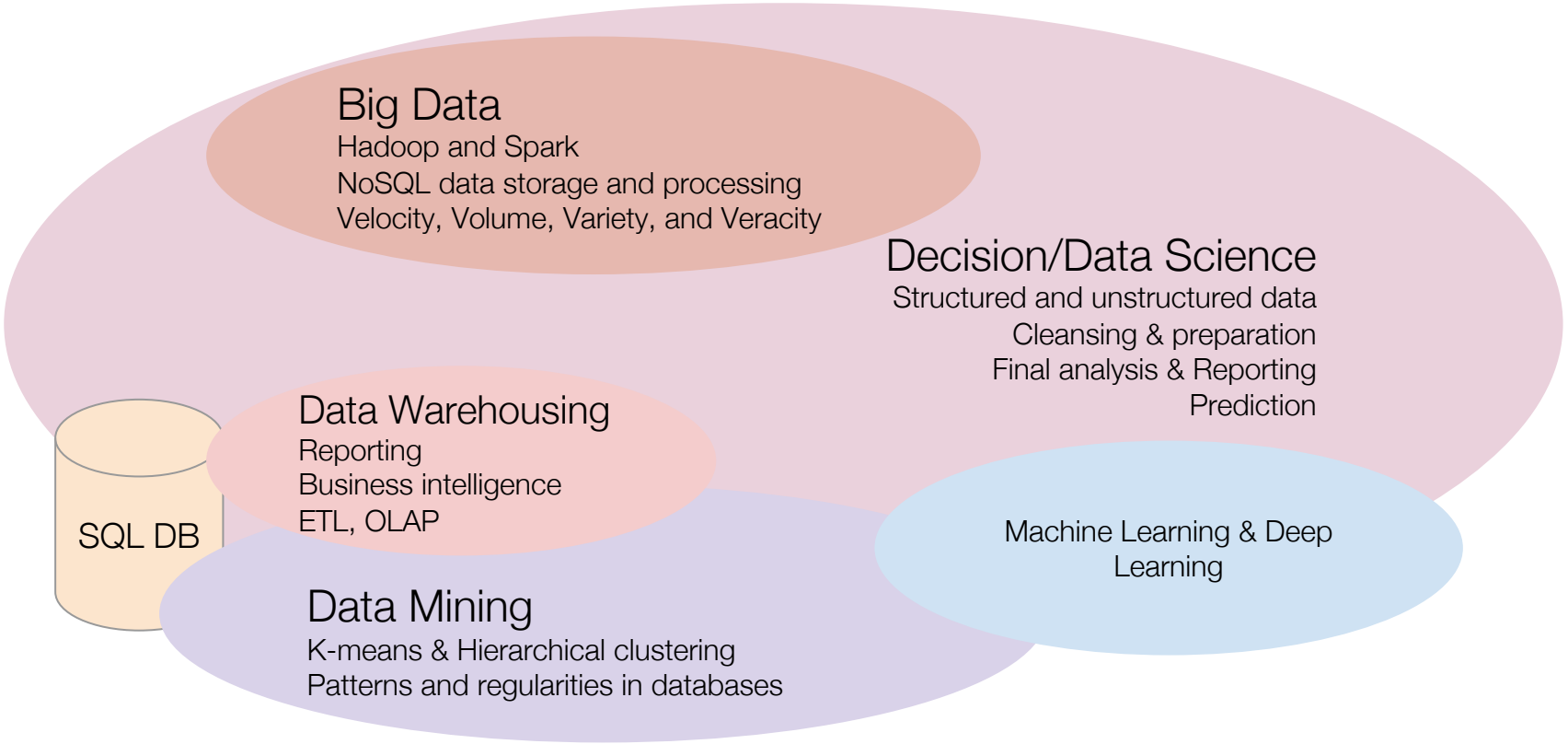
1980s

Deep Learning



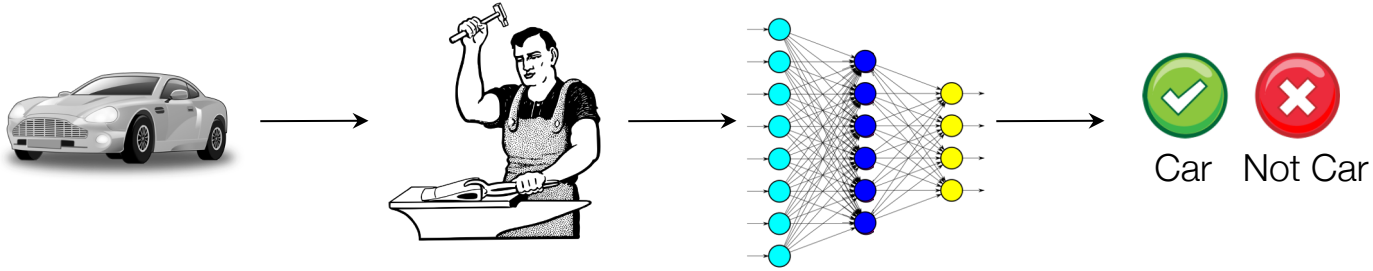
Powerful & trending
methods

2010s


Big Data
Hadoop and Spark
NoSQL data storage and processing
Velocity, Volume, Variety, and Veracity
Decision/Data Science
Structured and unstructured data
Cleansing & preparation
Final analysis & Reporting
Prediction
Data Warehousing
Reporting
Business intelligence
ETL, OLAP
SQL DB
Machine Learning & Deep Learning
Data Mining
K-means & Hierarchical clustering
Patterns and regularities in databases

1.1 ML & DL

Machine Learning



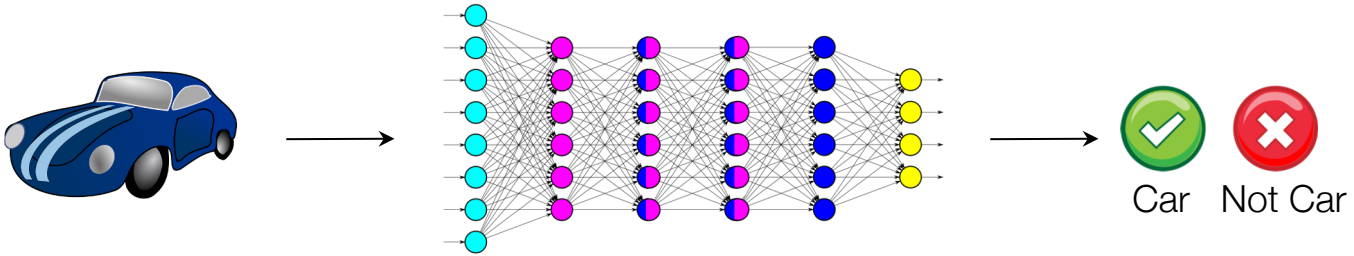
Input

Output

Feature Extraction

Classification

Deep Learning



Input

Output

Feature Extraction + Classification

Definition of machine learning - term coined in 1969 by Arthur Samuel

- ML is learning from examples “without being explicitly programmed”

- (some) ML Algorithms / Methods
  - Dimensionality reduction - e.g. Principal Component Analysis
  - Ensemble learning - e.g. Boosting using AdaBoost
  - Linear Classifiers - e.g. Logistic Regression
  - Supervised learning - e.g. Neural Networks and Support Vector Machines
  - Decision tree algorithms - e.g. Random Forest
  - Clustering - e.g. k-means algorithm
  - Unsupervised learning - e.g. Expectation-maximization algorithm
  - Deep Learning methods - e.g. Residual Networks
  - Bayesian networks - e.g. Naive Bayes classifier

XGBoost: A highly effective & widely used “ML” method

- XGBoost is an **open-source library**
- Is a **gradient boosting framework** for C++, Java, Python,R, and Julia
  - a prediction model in the form of an ensemble of weak prediction models (usually decision trees)
- Is popular among the Kaggle community
  - used for a large number of competitions

kaggle.com

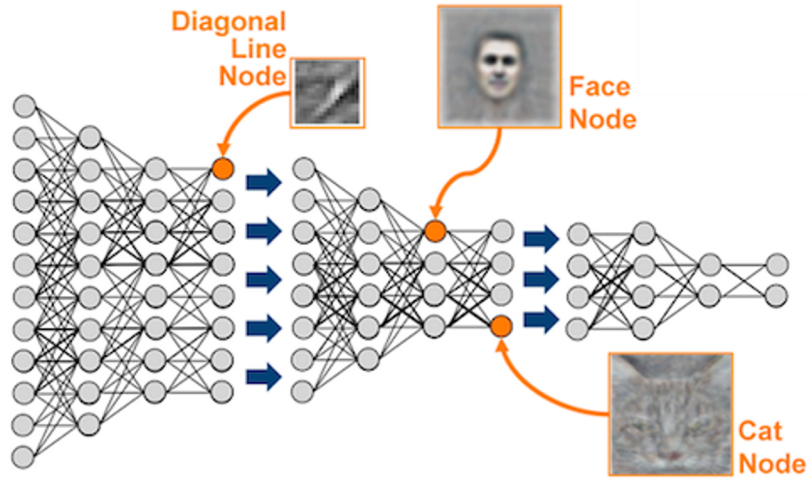
- Integrated with scikit-learn for Python, and with the caret package for R

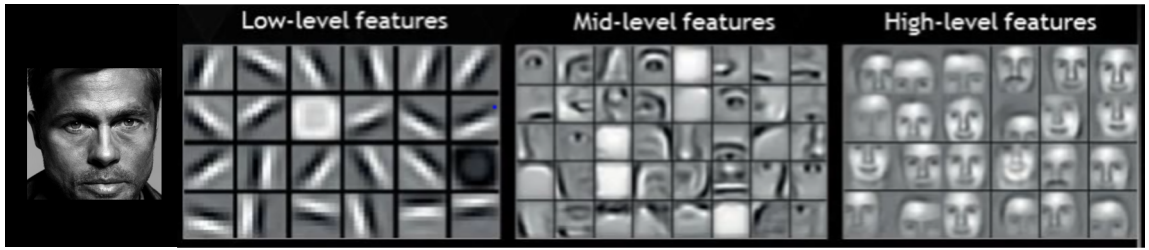
- Can be integrated into data flow frameworks like Apache Spark and Hadoop

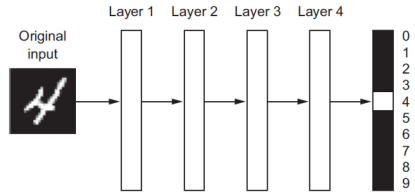
XGBoost: A Scalable Tree Boosting System

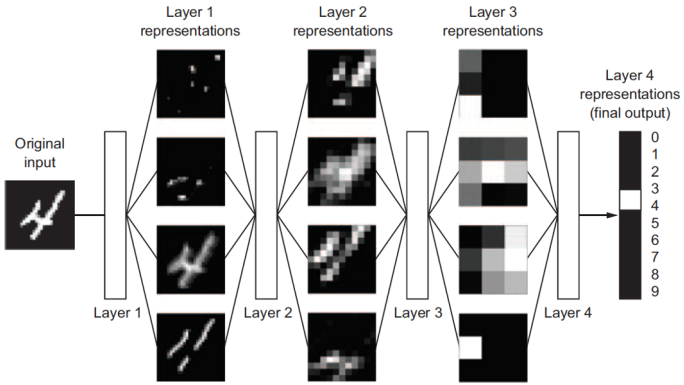
Tianqi Chen, Carlos Guestrin

1.1.4 DL is hierarchical feature learning



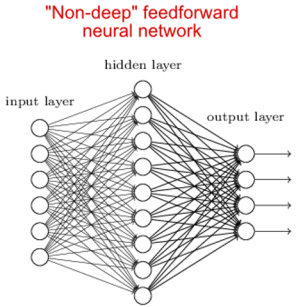


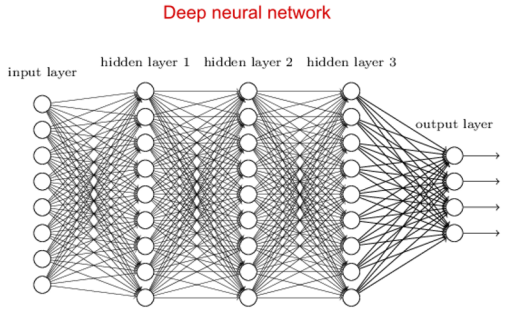




1.1.4 The “Deep” in deep learning

- Shallow Learning
  - ML methods that focus on learning only one or two layers of representations of the data
  - After two hidden layers, performance starts to drop
- Deep Learning
  - Modern deep learning often involves tens or even hundreds of successive layers of representations
    - they’re all learned automatically from exposure to training data
- Alternative names
  - layered representations learning and hierarchical representations learning
- Deep learning is a mathematical/computational framework for learning representations from data

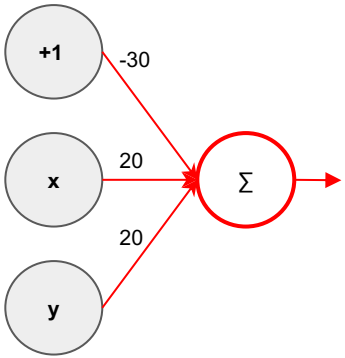




A hidden layer

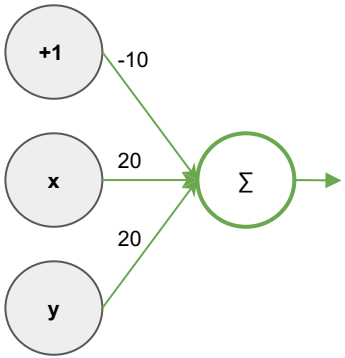
x AND Y

x	y	f-and(x,y)
0	0	0
0	1	0
1	0	0
1	1	1



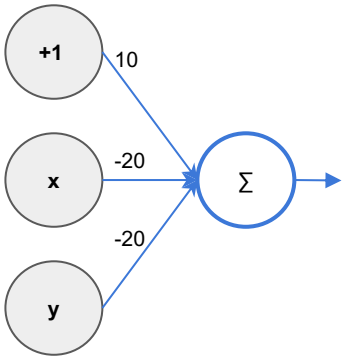
x OR y

x	y	f-or(x,y)
0	0	0
0	1	1
1	0	1
1	1	1



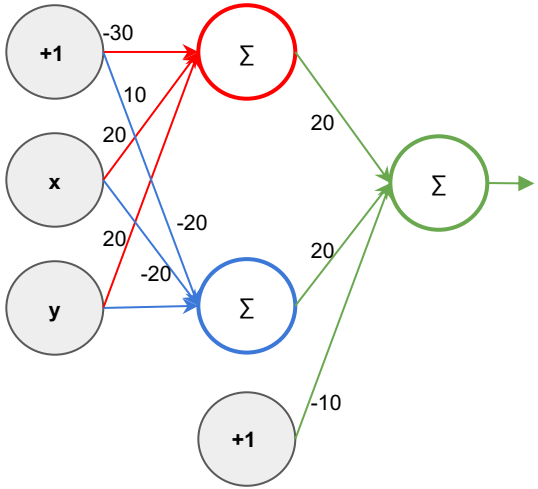
(!x) AND (!y)

x	y	f-rev-and(x,y)
0	0	1
0	1	0
1	0	0
1	1	0



x XNOR y

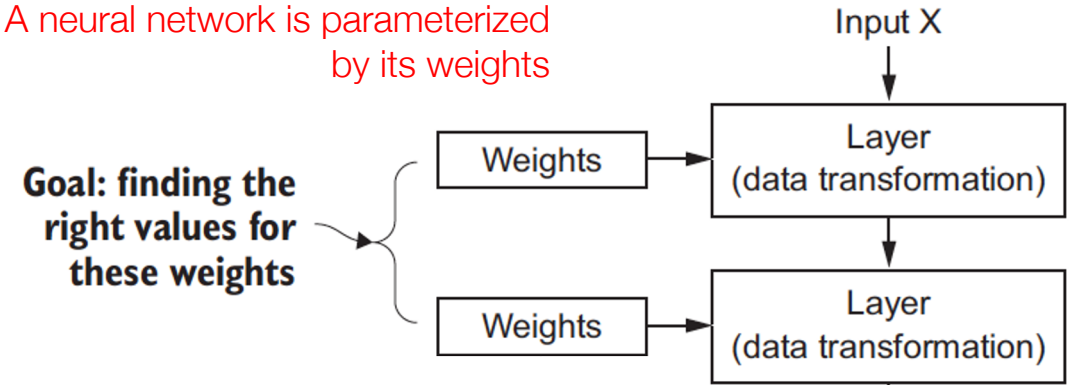
x	y	f-xnor(x,y)
0	0	1
0	1	0
1	0	0
1	1	1



XNOR = (a AND b) OR (!a AND !b)

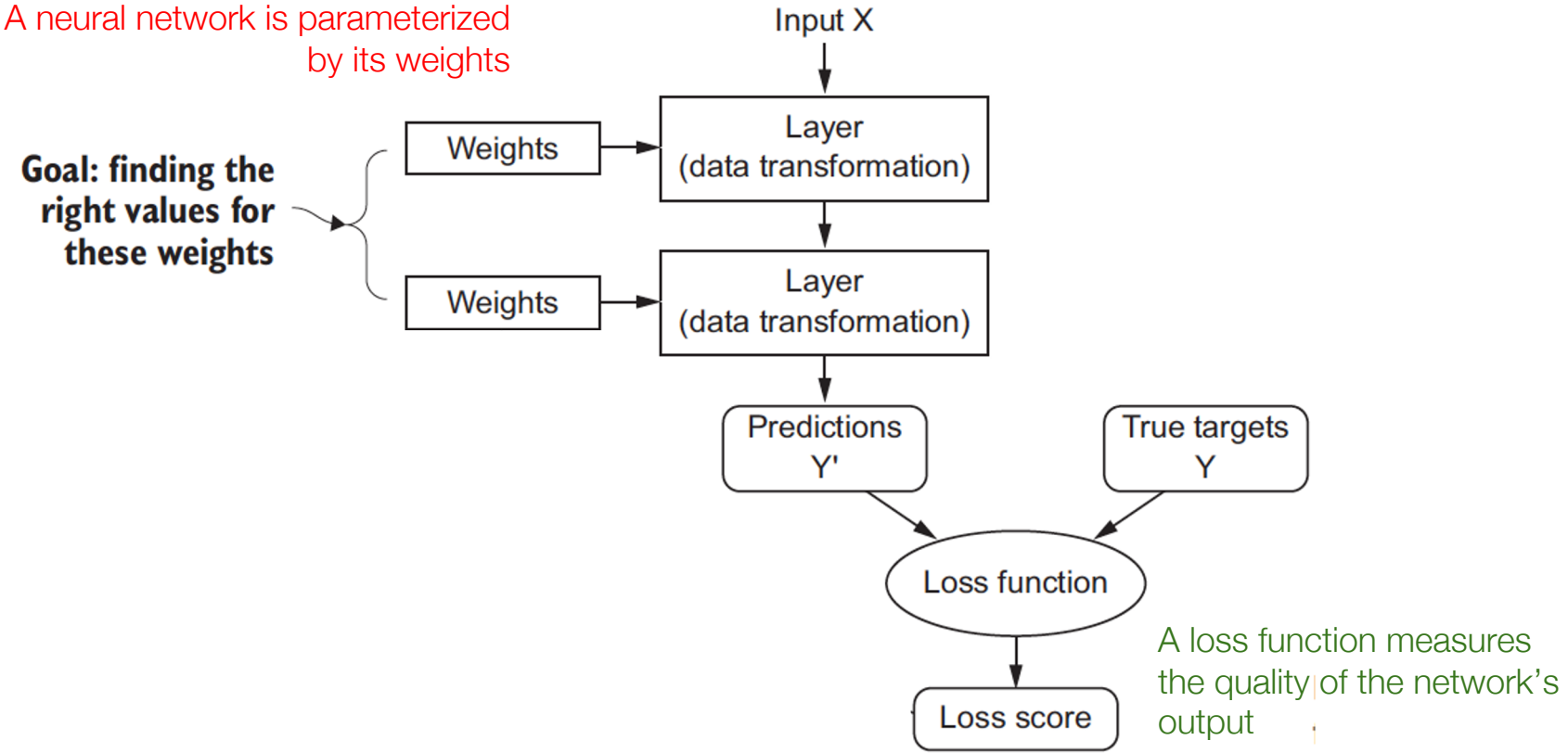
1.1.5 How does machine learning (or DL) work?

A neural network is parameterized by its weights



1.1.5 How does machine learning (or DL) work?

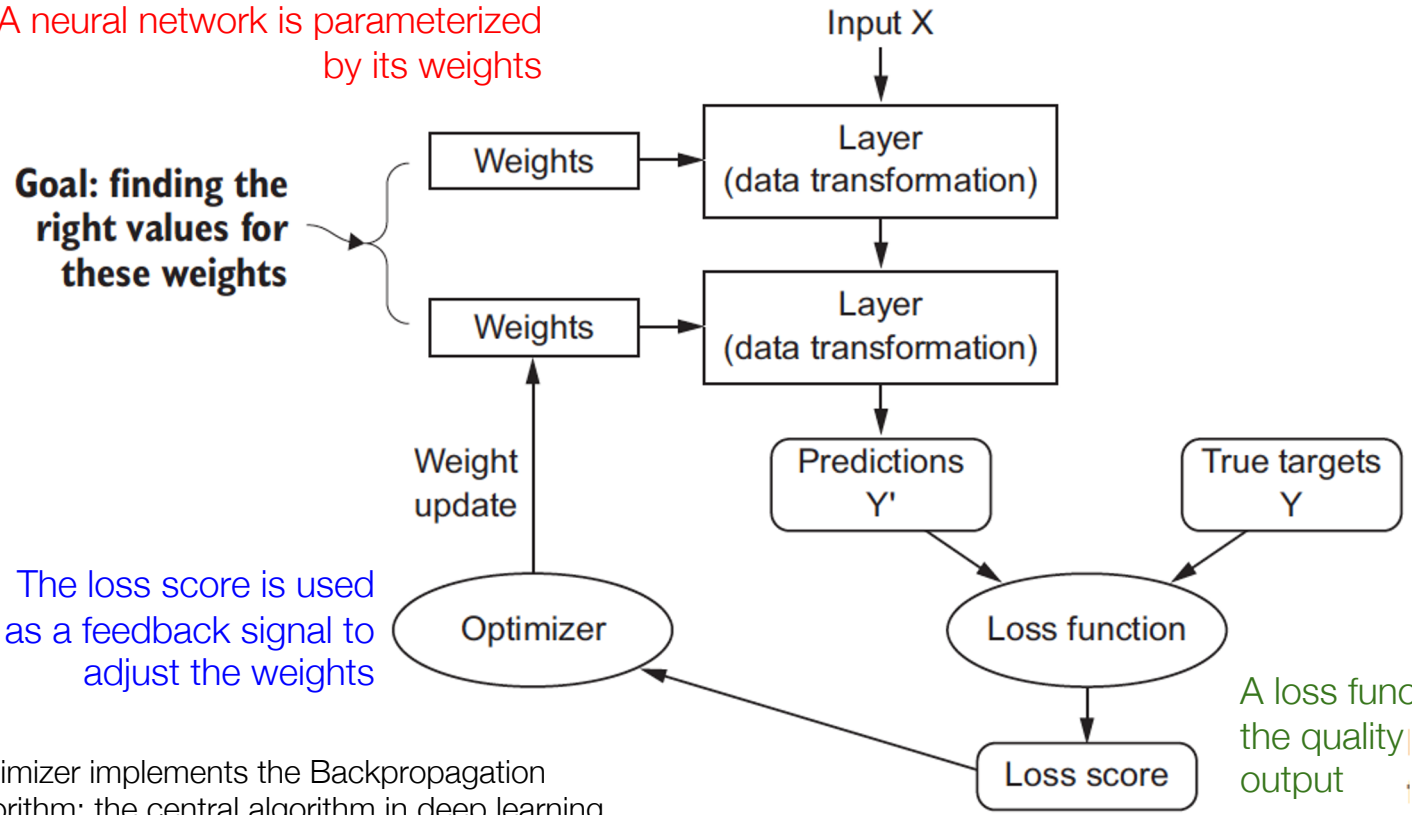
A neural network is parameterized by its weights



A loss function measures the quality of the network's output

1.1.5 How does machine learning (or DL) work?

A neural network is parameterized by its weights



The loss score is used as a feedback signal to adjust the weights

A loss function measures the quality of the network's output

Optimizer implements the Backpropagation algorithm: the central algorithm in deep learning

1.3 Why deep learning? Why now?

- The two key ideas of deep learning for computer vision—convolutional neural networks and backpropagation—were already well understood in 1989
- The Long Short-Term Memory (LSTM) algorithm, which is fundamental to deep learning for time series, was developed in 1997 and has barely changed since
- So why did deep learning only take off after 2012? What changed in these two decades?

1.3 Why deep learning? Why now?

- The two key ideas of deep learning for computer vision—convolutional neural networks and backpropagation—were already well understood in 1989
- The Long Short-Term Memory (LSTM) algorithm, which is fundamental to deep learning for time series, was developed in 1997 and has barely changed since
- So why did deep learning only take off after 2012? What changed in these two decades?
- In general, three technical forces are driving advances in machine learning:
  - Hardware (GPUs)
  - Datasets and benchmarks
    - ImageNet dataset - 1.4 million images hand annotated with 1,000 image categories
    - Kaggle competitions
  - Algorithmic advances
    - Better activation functions and better weight-initialization schemes
    - Better optimization schemes, such as RMSProp and Adam

http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

4.1 Four branches of machine learning

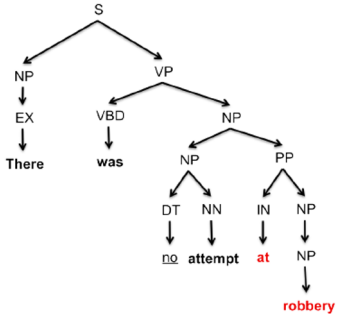
- Machine learning is a vast field with a complex subfield taxonomy
- Machine-learning algorithms generally fall into four broad categories:
  a) Supervised learning
  b) Unsupervised learning
  c) Self-supervised learning
  d) Reinforcement learning

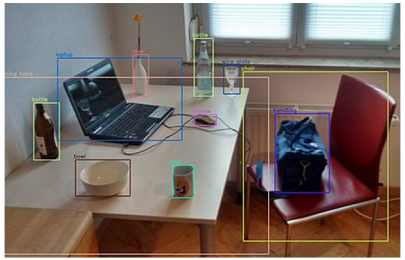
4.1.1 Supervised learning

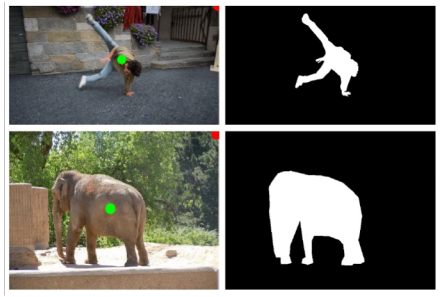
- It consists of learning to map input data to known targets (also called annotations), given a set of examples (often annotated by humans)
- Some ‘uncommon’ examples:
  - *Sequence generation*—Given a picture, predict a caption describing it
  - *Syntax tree prediction*—Given a sentence, predict its decomposition into a syntax tree
  - *Object detection*—Given a picture, draw a bounding box around certain objects inside the picture
  - *Image segmentation*—Given a picture, draw a pixel-level mask on a specific object

A female tennis player in action on the court.

A group of young men playing a game of soccer







https://machinelearningmastery.com/how-to-caption-photos-with-deep-learning/

4.1.2 Unsupervised learning

- Finding interesting transformations of the input data without the help of any targets
  - For the purposes of (a) data visualization, (b) data compression, (c) data denoising, or (d) to better understand the correlations present in the data at hand
- Unsupervised learning is the bread and butter of data analytics
  - It's often a necessary step in better understanding a dataset before attempting to solve a supervised-learning problem (basic NN training is slower when two features are highly correlated)
- "Dimensionality reduction" and "clustering" are well-known categories of unsupervised learning, e.g. Principal component analysis

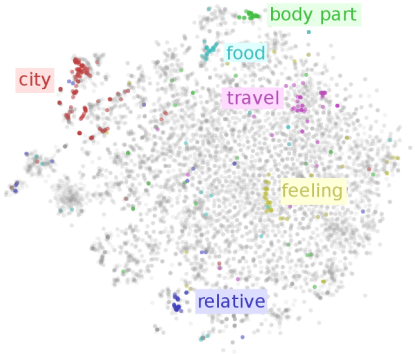


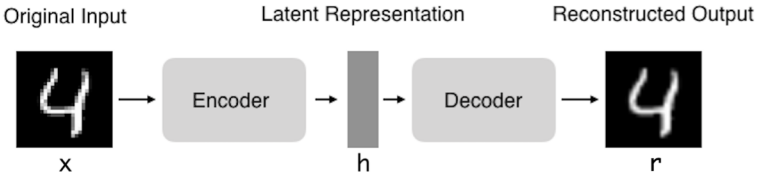
Image Restoration - https://en.wikipedia.org/wiki/Image_restoration

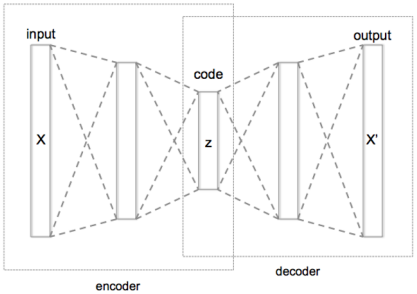
PCA - https://www.youtube.com/watch?v=HMOI_lkzW08

Correlation vs Causation - https://towardsdatascience.com/why-correlation-does-not-imply-causation-5b99790df07e

4.1.3 Self-supervised learning

- Self-supervised learning is supervised learning without human-annotated label
  - supervised learning without any humans in the loop
- There are still labels involved
  - they’re generated from the input data, typically using a heuristic algorithm
- “autoencoders” are an instance of self-supervised learning





- Other examples:
  - Trying to predict the next frame in a video, given past frames
  - Trying to predict next word in a text, given previous words

4.1.3 Self-supervised learning

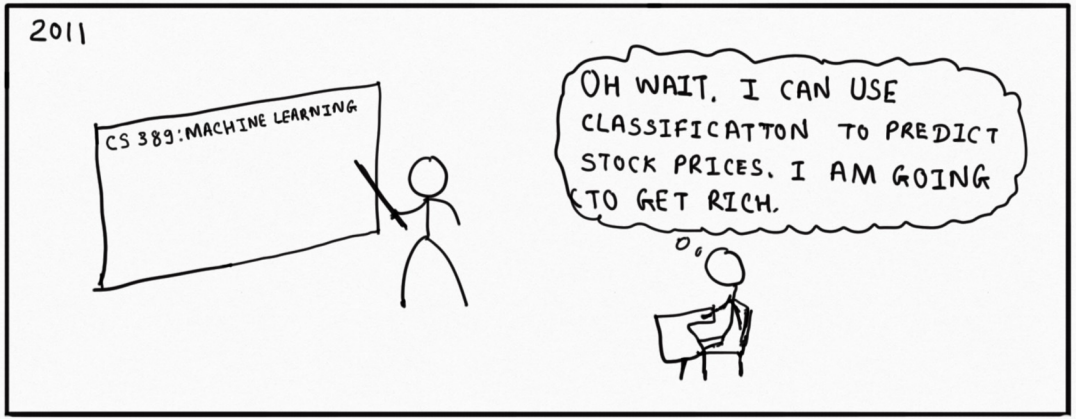
- The distinction between supervised, self-supervised, and unsupervised learning can be blurry sometimes
  - These categories are more of a continuum without solid borders
  - Self-supervised learning can be reinterpreted as either supervised or unsupervised learning, depending on whether you pay attention to the learning mechanism or to the context of its application

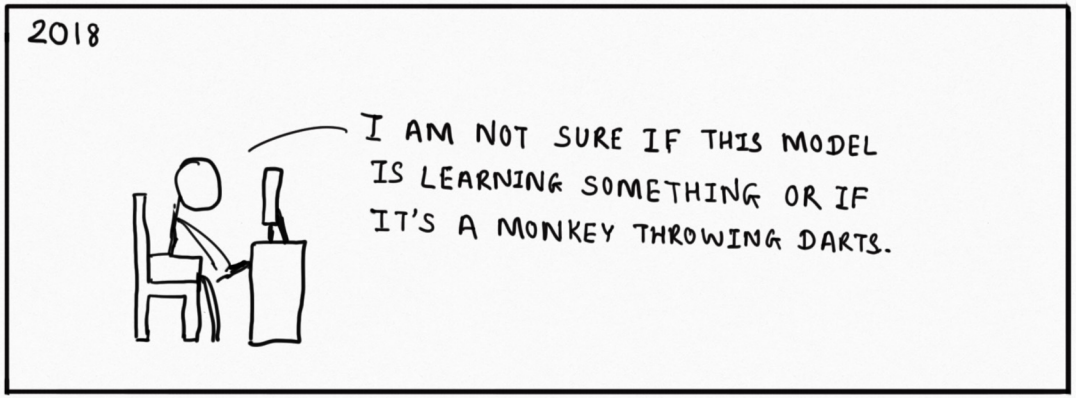
4.1.4 Reinforcement learning

- Reinforcement learning is mostly a research area and hasn't yet had significant practical successes beyond games
- In reinforcement learning, an agent receives information about its environment and learns to choose actions that will maximize some reward
  - For instance, a neural network that "looks" at a videogame screen and outputs game actions in order to maximize its score can be trained via reinforcement learning

Learning ‘bleeding-edge’ Deep Learning

because deep learning is already cutting-edge





Learning ‘bleeding-edge’ Deep Learning because deep learning is already cutting-edge

Approach 1:

Learn basic mathematics in machine learning

then, learn basics of machine learning

then, learn basic concepts of deep learning

then, practice deep learning

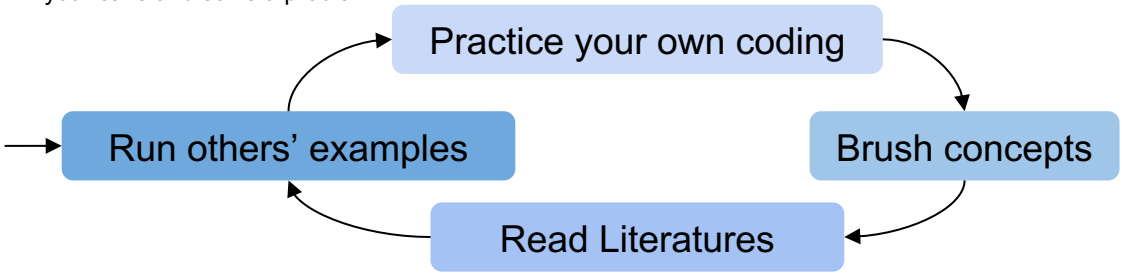
Approach 2:

Pick a project

then, learn everything you need to learn to solve it

Approach 3:

Theory and application is tied together very closely - you could not just study the theory and go to your cave and solve a problem



The history of deep learning