Machine Learning 101

Catherine Zeng, Dell EMC Intern Advanced Server Engineering
Studying EECS, Massachusetts Institute of Technology '20

It's an exciting time for ML. There is ample work to be done at all levels: from the theory end to the framework end, much can be improved. It's almost as exciting as the creation of the internet.

- Ryan Dahl, author of Node.js

What is Machine Learning?

Machine learning creates models that are built automatically using algorithms that iteratively learn from data, so that the machine can find patterns in data without the need for us to define them.

Where have you seen Machine Learning?

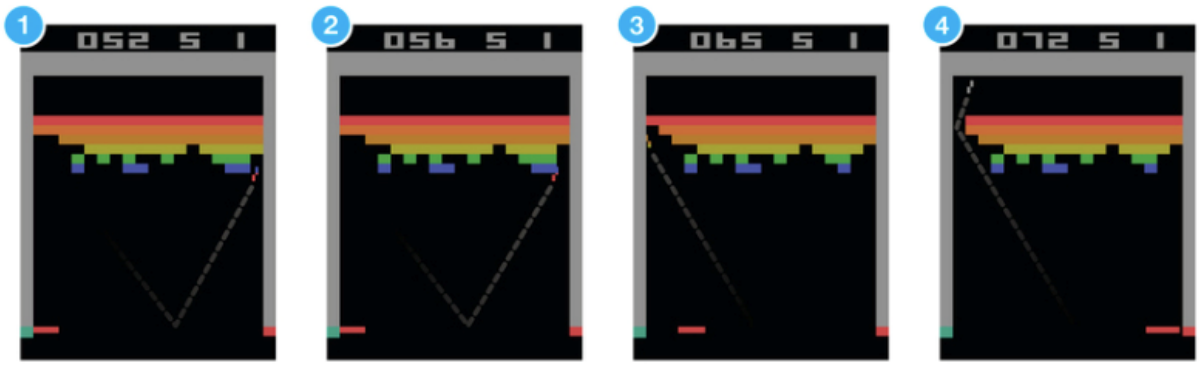
Recommender problems: Amazon or Netflix analyzing your activity and comparing it to other users with similar activity to make recommendations.

Game Playing: Using deep neural networks and tree search, AlphaGo is a Machine Learning powered computer program that beat the world's No.1 ranked player Ke Jie.

Self-Driving Cars: Google, Apple, and Tesla, among most other car manufacturers, are developing autonomous vehicles that are trained using computer vision.

Computer-aided detection (CADe) and diagnosis (CAD): Support vector machines (SVMs), linear discriminant analysis, and deep neural networks, among other algorithms, are being used in medical imaging to diagnosis.

QUESTION: Can you think of Machine Learning in your life? (Type response)



Link to Google DeepMind's Deep Q-Learning playing Atari Breakout here.

Stages of Machine Learning Implementation:

1) Determining objectives and metric: what is the purpose of this model? How will the quality of the model be assessed?
2) Preparing the dataset:
    a. Select the data relevant to the question or problem and evaluate the quality of the dataset.
    b. Preprocess the data, which includes formatting, cleaning data to fill in missing values, or taking a smaller sample of the data.
    c. Transform data using knowledge of the algorithm you are working with and problem domain. This includes scaling data, splitting data, or aggregating data.
3) Model training: Select a machine learning algorithm to apply to the data, with will learn patterns and output a model that matches these relationships.
4) Integration and testing: Evaluate the quality of the model on test data, and continue monitoring the quality of the model on new testing data.

QUESTION: What is of the following is NOT considered steps to preprocess data? A) Taking a smaller sample of the data, B) Formatting the data, C) Making up more data for a larger sample size, or D) Cleaning data to fill in missing values.

ANSWER: C) Making up more data for a larger sample size. From the excerpt: "Preprocess the data, which includes formatting, cleaning data to fill in missing values, or taking a smaller sample of the data."

Key Terms

Training set: dataset used to find relationships between data

Testing set: Data used to evaluate the quality of the trained model.

Training Error: Error when you run the trained model on the training set.

Testing Error: Error when you run the trained model on the testing set, which the model hasn't seen yet.

QUESTION: Which do you think would be higher, A) the training error or B) the testing error?

ANSWER: The testing error would be higher because the model is applied to data it has never seen before, whereas for the training error the model is applied to the very data it was built off of!

Machine Learning algorithms:

Supervised Learning:

> The system is given a training set that is labeled, and asked to label a testing set. Examples: Linear classifiers, Perceptron algorithm, Support Vector Machines, Neural Networks, Naïve Bayes. Supervised learning often involves classification and regression. Classification separates data into distinct classes and include SVMs and decision trees.

Unsupervised Learning:

> The system is given unlabeled data and asked to classify the data. Examples: Gaussian Mixture Models, K-Means algorithm. Unsupervised learning often includes clustering, which groups data together and finds similarities between them.

Reinforcement Learning:

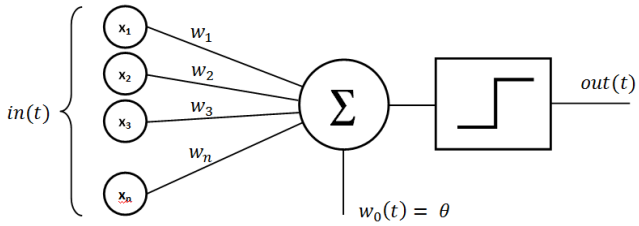
> The system learns actions to maximize rewards in a reward system. Examples include Q-learning which can be used to find optimal policies for Markov decision processes. Reinforcement learning is used to teach computers to play video games. The most famous is Google DeepMind's deep reinforcement learning model playing Atari.

QUESTION: Which if the following is not considered supervised learning? You can select more than one. A) Support Vector Machines, B) Gaussian Mixture Models, C) the Perceptron Algorithm, D) Q-Learning.

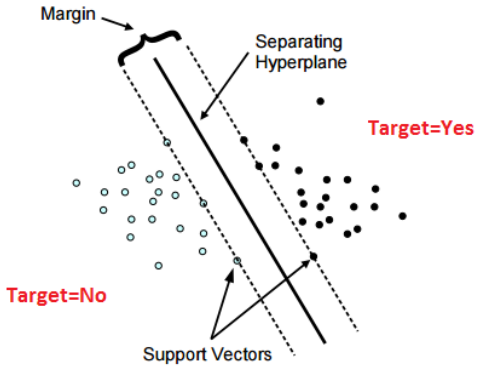
ANSWER: B) Gaussian Mixture Models, and D) Q-Learning. Gaussian Mixture Models is unsupervised learning, and Q-Learning is reinforcement learning.

Overview of ML Algorithms and Important Concepts

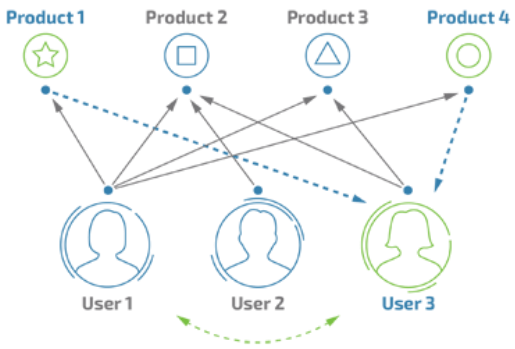
Perceptron – The perceptron is the most basic machine learning algorithm that learns a linear classifier to separate two groups of data.



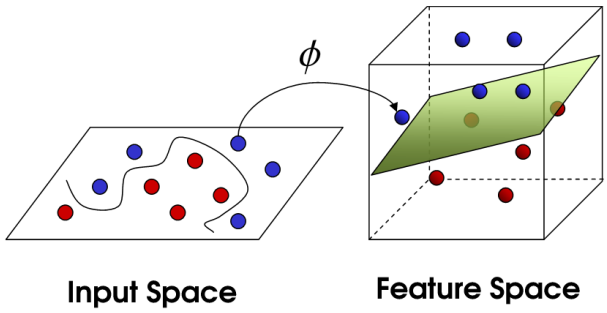
Support Vector Machine (SVM) – A linear classifier that separates two groups of data through a separating hyperplane.



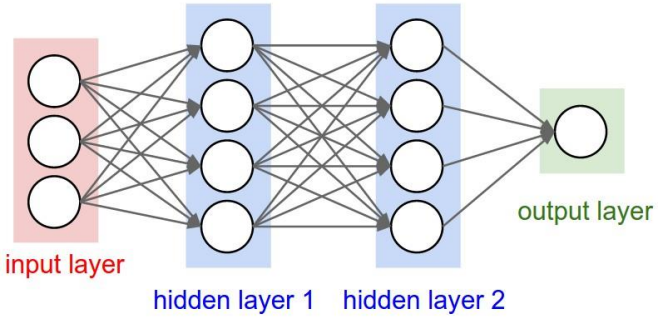
Collaborative filtering and Recommender systems: Makes predictions or recommendations about a user based on other users in the system.



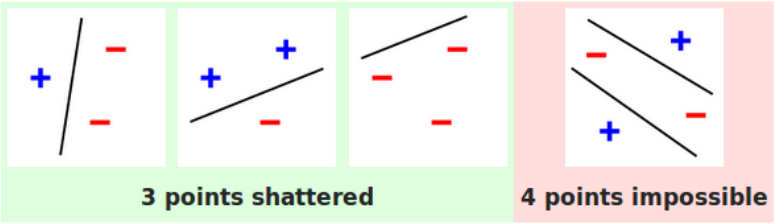
Kernels: Makes nonlinear data compatible with linear models by mapping data to higher dimensions that show linear patterns.



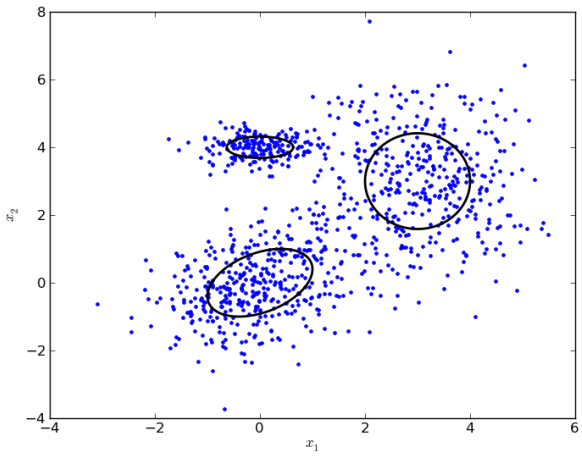
Neural Networks: Made up of layers of interconnected nodes that contain an activation function. Modeled after our understanding of biological neurons. Popular neural networks include recurrent neural networks and convolutional neural networks.



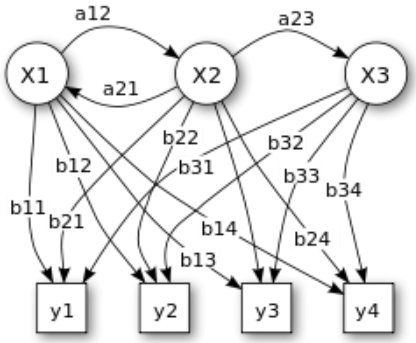
Vapnik–Chervonenkis theory (VC Dimension): a measure of the capacity of a classifier, which includes complexity, expressive power, richness, or flexibility.



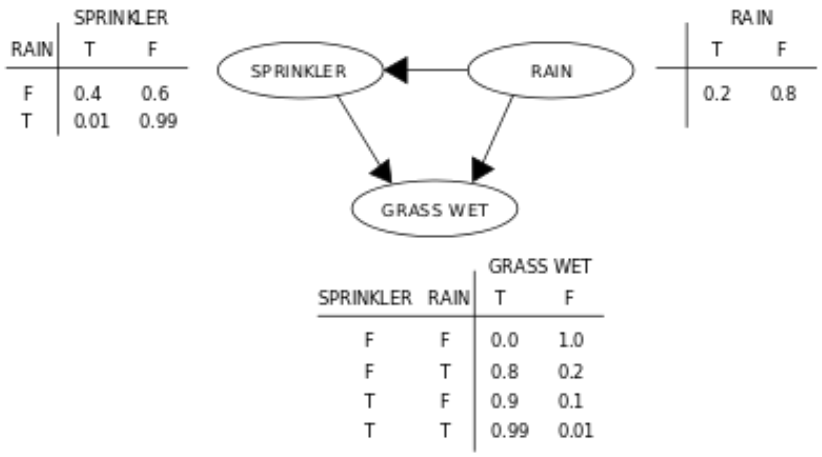
Gaussian Mixture Models: probabilistic models to represent normally distributed clustering of data; a type of unsupervised learning.



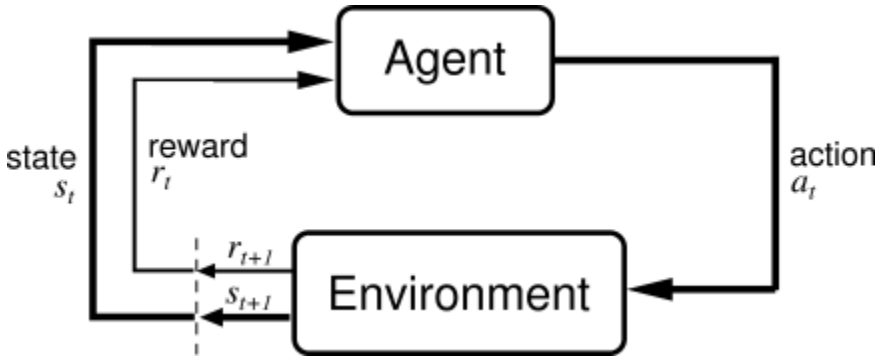
Hidden Markov Models (HMMs): HMMs are a set of states that are each associated with a probability distribution.



Bayesian network: a probabilistic model that represents random variables and their conditional dependencies.



Reinforcement learning: how to take actions in an environment to maximize reward with respect to a predetermined goal.

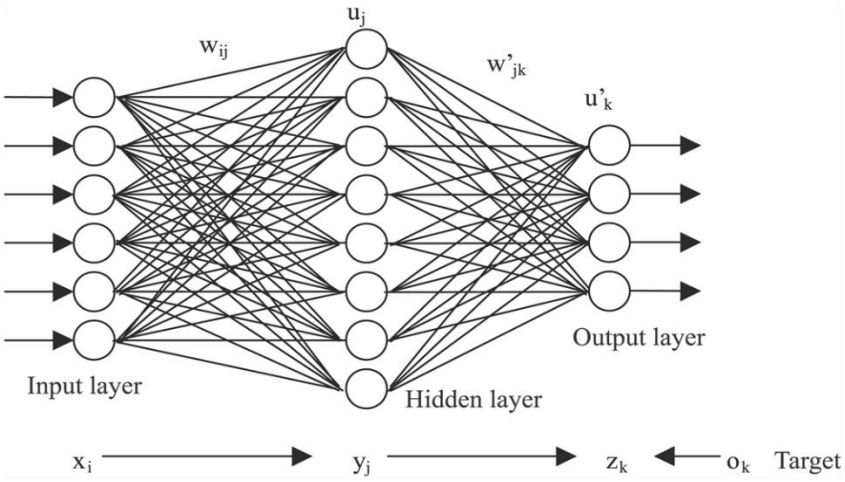


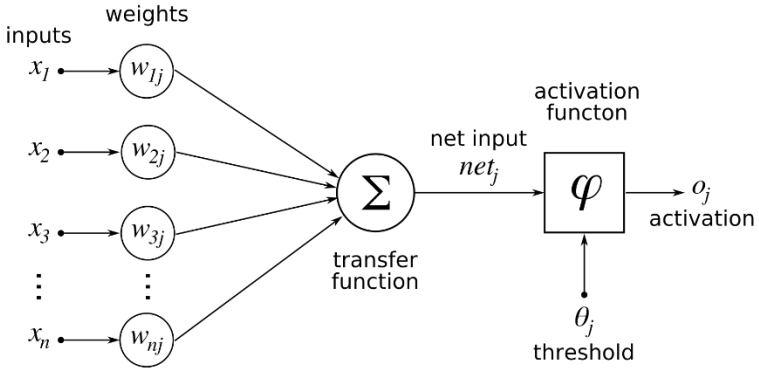
QUESTION: What is your favorite machine learning algorithm?

Artificial Neural Networks (ANNs)

ANNs are based on our understanding of biological neural networks, and is based on the basic perceptron. ANNs contain an input layer, a hidden layer, and an output layer. It is made of interconnected nodes with weights and an activation functions that describe the strength of the connection.

For more information, read this.

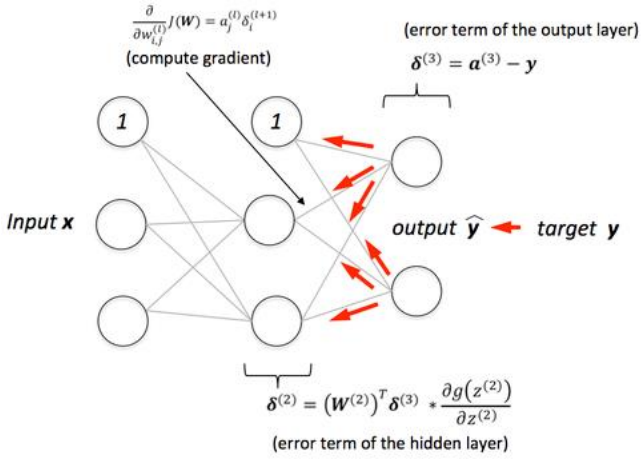




QUESTION: What is the basic linear classified that artificial neural networks are based off of? A) Bayes Tree, B) Naïve Bayes, C) Support Vector Machines, D) the Perceptron

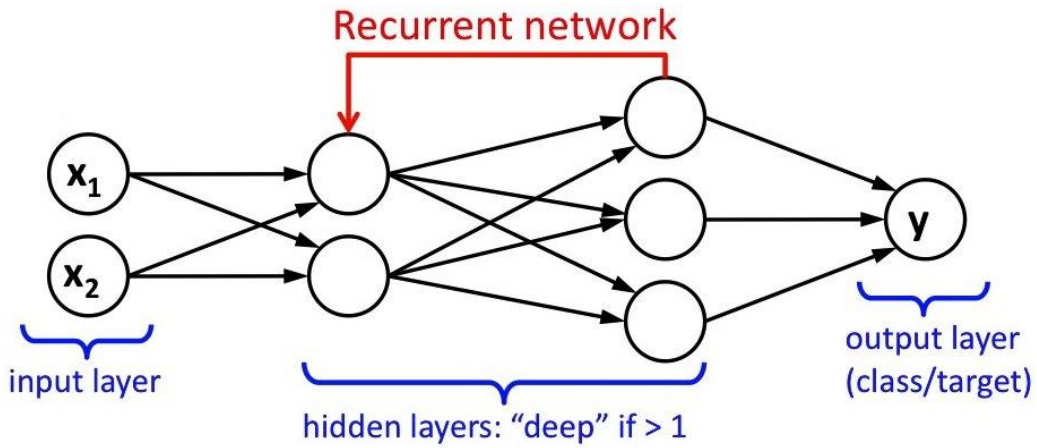
ANSWER: D) the Perceptron

Backpropagation is used in ANNs to calculate the gradient of the loss function with respect to weights. Backpropagation better adjusts weights in a neural network model so that it has better performance.

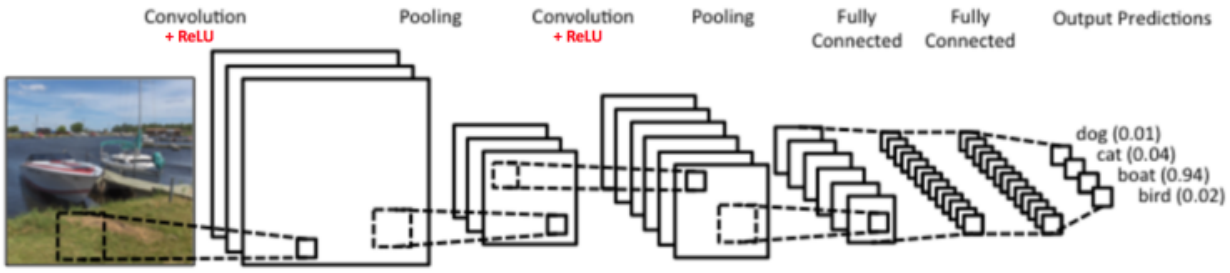


The feedforward neural network is the simplest class of artificial neural networks. Information in the feedforward neural network only moves in the forward directly. Other common classes of artificial neural networks are the **recurrent neural network** and **convolutional neural network.**

Recurrent Neural Network (RNN): Unlike the feedforward neural network, information moves in a cycle in the recurrent neural network. This means that the RNN depends on the output of the previous step, which is analogous of a "memory" that stores previous outputs. RNNs are arguably the most powerful type of neural network and is commonly used in natural language processing tasks.



Convolutional Neural Network (CNN or ConvNet): CNNs are primarily used for visual imagery and is common in computer vision tasks. CNNs are designed to deal with 2D input structures and consists of three types of layers: the convolutional layer, pooling layer, and fully-connected layer. In 2015, a CNN was able to beat a human in object recognition for the first time.



Artificial neural networks that contain more than one hidden layer are considered to be **deep learning**. Deep learning is part of the learning data representations method of machine learning, which is based on feature detection, and can range anywhere from supervised to unsupervised learning. Because deep learning is so computationally expensive, they have popularized the usage of **graphics processing units (GPUs)** because of their **higher memory bandwidth**.

QUESTION: How many layers does deep learning have? A) 1, B) 2, C) 3 D) 4

ANSWER: Trick question! Deep learning is considered to be artificial neural networks with more than one hidden layer in addition to the input and output layer. So Deep Learning is considered to be ANNs with more than three layers.

QUESTION: Which of the following algorithms is used primarily in computer vision? A) Convolutional Neural Networks, B) Backpropagation, C) Recurrent Neural Networks, D) GPUs

ANSWER: A) Convolutional Neural Networks. ConvNets are primarily used for visual imagery and is common in computer vision tasks.

QUESTION: Why are GPUs faster than CPUs? A) They are more expensive, B) They contain deep learning inspired hardware, C) They consume more power than GPUs do, D) They have a higher memory bandwidth

ANSWER: D) They have a higher memory bandwidth. The next section will explain this more in detail.

GPUs and Nvidia

Memory bandwidth is the rate at which data can be read from or store into a semiconductor memory by a processor. GPUs have much higher memory bandwidths than CPUs do, making GPUs better at doing matrix multiplication and convolution common in deep learning. GPUs can have up to 750 GB/s memory bandwidth while CPUs can only have up to 50 GB/s.

GPUs have higher latency than CPUs do, but this is masked under thread parallelism. Overall, GPUs are MUCH faster at processing large amounts of data than CPUs are. For an analogy, see here.

At the moment, **Nvidia** dominates the market for GPUs with models ranging from its **Titan X** to its **Tesla P4**, which are compatible for usage with Dell's servers.

QUESTION: True or False, CPUs have a better memory bandwidth than GPUs do.

ANWER: False. GPUs are able to better run machine learning tasks because it has a higher memory bandwidth.

Important People in Machine Learning Development

Geoffrey Hinton, professor at the University of Toronto and Engineering Fellow at Google managing the Brain Team Toronto. Hinton is the most important figure in deep learning. He demonstrated backpropagation for deep neural nets, co-invented Boltzmann machines (a type of recurrent neural network), and more.

Yann LeCun, Director of AI Resarch at Facebook and professor at NYU's Courant Institute of Mathematical Sciences. LeCun is known for his work using convolutional neural networks for computer vision and is known as the founding father of CNNs.

Andrew Ng, former chief scientist at Baidu, adjunct professor at Stanford University, and co-founder of Coursera. Ng is known for his popular Machine Learning course on coursera and for popularizing machine learning concepts with the general public.

Peter Norvig, Director of Research at Google. Norvig is known for being the c-author of the leading textbook *Artificial Intelligence: A Modern Approach*

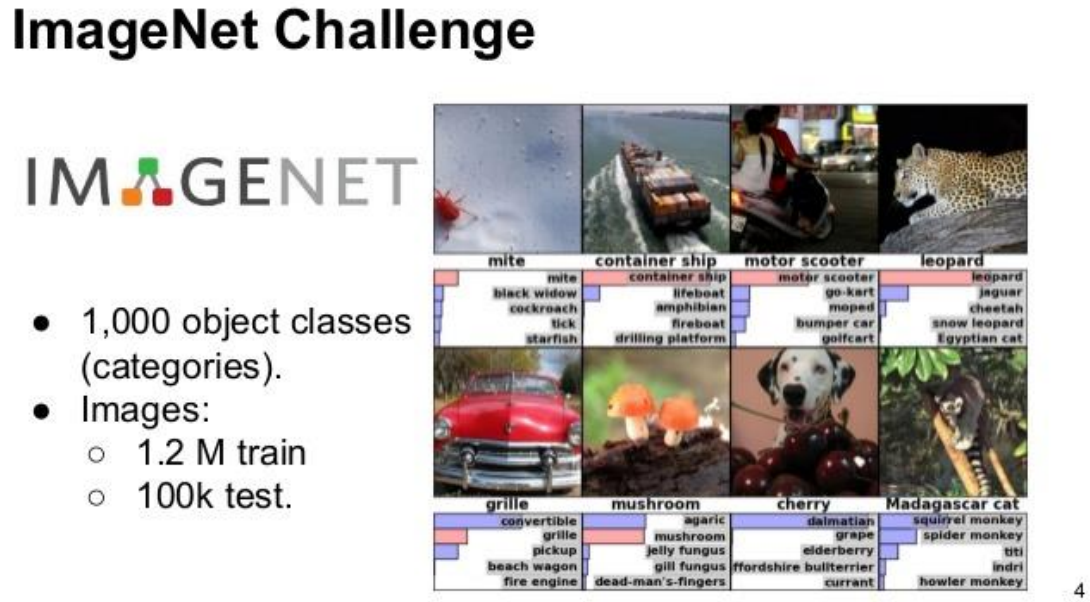
Michael I. Jordan is a professor at the University of California, Berkeley, and is known for popularizing Bayesian networks and for his research in the intersection of Machine Learning and statistics.

QUESTION: Who is known as the father of deep learning? A) Yan LeCun, B) Peter Norvig, C) Michael Jordan, D) Geoffrey Hinton

ANSWER: D) Geoffrey Hinton. Hinton is well-known as the father of deep learning.

ImageNet Project and Challenge

The ImageNet project is a database of annotated images used for research in visual object recognition developed by Stanford professor Fei-Fei Li. Every year since 2010, ImageNet has ran the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) for contenders to develop algorithms that most accurately do image recognition on the ImageNet dataset. The most revolutionary winners of the ImageNet challenge include AlexNet (2012) and GoogLeNet (2014).



For More Resources

Follow Andrew Ng's popular Machine Learning course on Coursera here.

Follow Geoffrey Hinton's Machine Learning course on Coursera here.

Synopsis of some Machine Learning papers here.

Popular YouTuber Siraj Raval explains Machine Learning concepts here.

Recently appointed Director of AI at Tesla Andrej Karpathy's blog here.

Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville, the most updated and popular deep learning textbook here.

Conferences to attend:

Academic:

NIPS (Neural Information Processing Systems)

ICML (International Conference on Machine Learning)

KDD (Knowledge Discovery and Data Mining)

ICLR (International Conference on Learning Representations)

ACL (Association for Computational Linguistics)

EMNLP (Empirical Methods in Natural Language Processing)

CVPR (Computer Vision and Pattern Recognition)

ICCF (International Conference on Computer Vision)

Professional:

O'Reilly Artificial Intelligence Conference

Machine Learning Conference (MLConf)

AI Expo (North America, Europe, World)

AI Summit

AI Conference