Artificial Intelligence, Machine Learning and Deep Learning

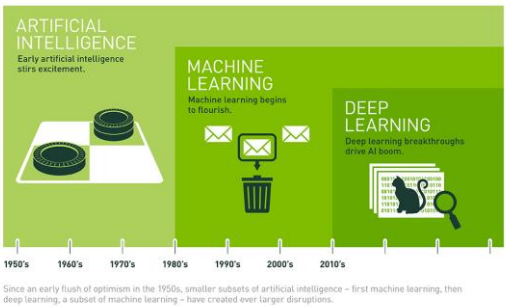
2020 - 2021

Ando Ki, Ph.D.
adki@future-ds.com

Contents

- AI, ML, and DL
- Terminologies
  - Regression
  - Classification
  - Clustering
- Types of learning of machine learning
- Data sets
- Training strategy
- Deep learning design flow
- Popular deep neural networks
- Data sets
  - MNIST
  - CIFAR-10
  - ImageNet

AI, ML, and DL



Deep Learning is a domain of Machine Learning and they are state-of-the-art approaches of AI (source: NVIDIA Blog)

딥러닝은 머신러닝의 한 방법이며, 이들은 인공지능의 최신기술이다.

- Artificial Intelligence — Human intelligence exhibited by machines
  - the capability of a machine to imitate intelligent human behavior
- Machine Learning — An approach to achieve artificial intelligence
- Deep Learning — A technique for implementing machine learning based on artificial deep neural network

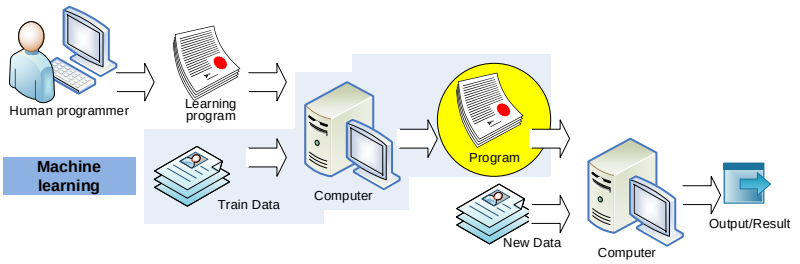
Machine learning (ML) is a subset of AI.
Deep learning (DL) is a subset of ML.

https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

ML

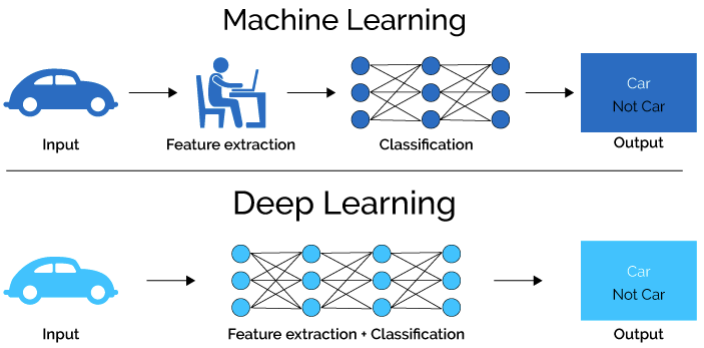
- Machine learning (ML: 기계학습)
  - ML gives computers the abilities to learn without being explicitly programmed to complete a task. – Arthur Samuel, 1959.
    - (기계학습은 컴퓨터가 '특정한 일을 처리하도록' '명시적인 프로그래밍 없이' '배울 수 있게하는 능력')

기계학습이란 "컴퓨터에 명시적으로 프로그래밍하지 않고 학습 할 수 있는 능력을 부여하는" 것

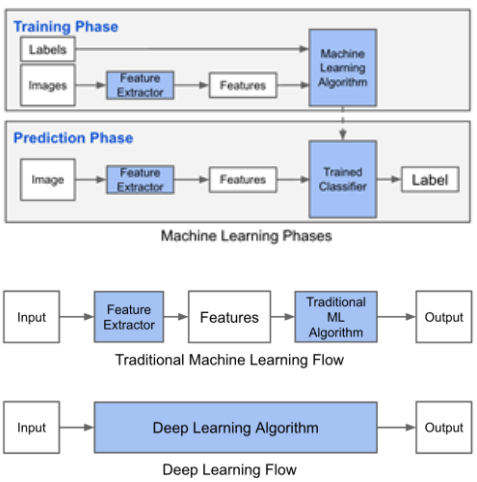


DL

- Deep learning (DL: 딥러닝)
  - A technique for implementing machine learning based on artificial deep neural network
  - (심층 인공 신경망을 이용한 기계학습의 한 구현 방법)



ML v.s. DL



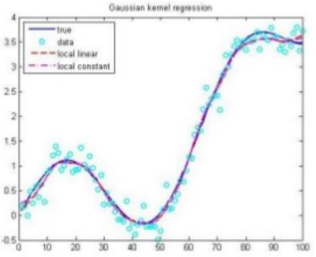
- Training phase: train a machine learning algorithm using a dataset comprised of the images and their corresponding labels.
  - Feature Extraction: utilize domain knowledge to extract new features that will be used by the machine learning algorithm.
  - Model Training: utilize a clean dataset composed of the images' features and the corresponding labels to train the machine learning model.
- Prediction phase: utilize the trained model to predict labels of unseen images.
- Differences between ML v.s. DL
  - ML: hand-craft feature engineering
  - DL: automatic feature engineering by the algorithm

Terminologies

- Regression (회기, 回歸)
  - A statistical technique used to measure the relation between variables (변수들 사이의 관계를 유추하는 통계적 기법)
  - predict new values based on the past, inference
    - an attempt to predict a continuous attribute
  - compute the new values for a dependent variable based on the values of one or more measured attributes (한 개 또는 여러 개의 측정된 특성으로 종속변수의 새 값을 계산)
- Clustering (군집화, 群集化)
  - partitioning of a data set into subsets (clusters) so that data in each subset ideally share some common characteristics (공통된 특징을 갖는 보다 작은 집합으로 나누는 것)
  - 입력 데이터의 분포 특성(입력값의 유사성)을 분석하여 임의의 복수 개의 그룹으로 나누는 것
- Classification (분류, 分類)
  - divide samples in classes (샘플을 클라스로 나눈다)
  - use a trained set of previously labeled data (미리 꼬리표(label)가 붙은 샘플로 트레이닝한다.)
  - It is similar to the clustering, but requires that the analyst knows ahead of time how classes are defined. (이것은 군집화와 유사하지만, 미리 클라스가 정의되어 있다는 점에서 차이가 있다.)
  - 주어진 데이터 집합을 이미 정의된 몇 개의 클래스로 구분하는 것

Terminologies

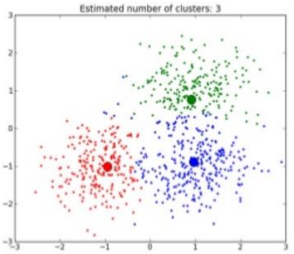
regression
real target values
x ∈ (0,100)
y : real number



classification
discrete target values
x : pixels (28*28)
y : 0,1, 2,3,...,9

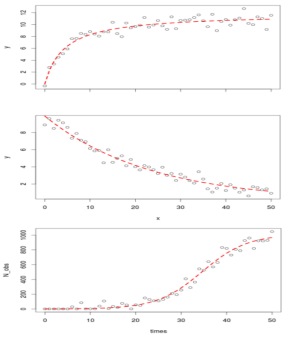
Label A ➔ means '8'

clustering
no target values
x ∈ (-3,3)×(-3,3)

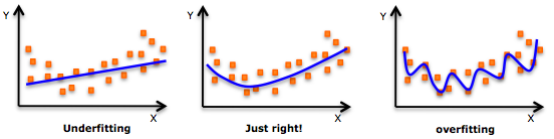


Regression

- Curve fitting
  - find a well defined and known function underlying data (주어진 데이터에 잘 맞는 알려진 함수를 찾는 것)
    - **Linear regression**

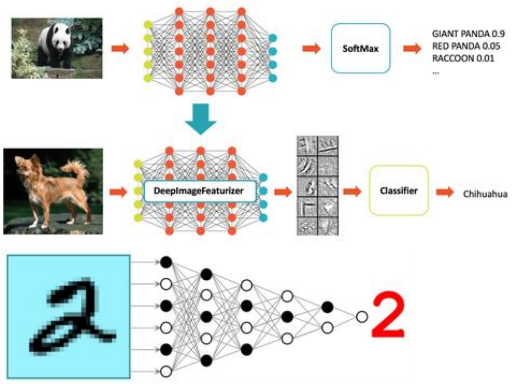


- Data table statistical correlation
  - mapping without any prior assumption on the functional form of the data distribution
    - **Logistic regression**
  - machine learning algorithms well suited for this.



Classification

- To predict group membership for data instances
  - Probabilistic classification (통계적 분류)
    - given an input, the classifier returns its probabilities to belong to each class
  - Crispy classification (결정적 분류)
    - given an input, the classifier returns its label

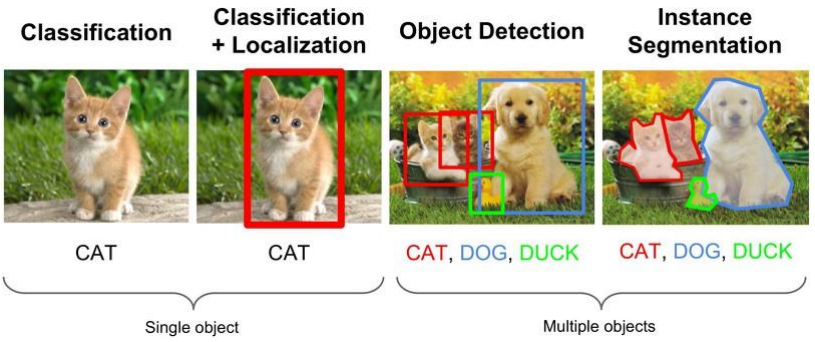


Softmax is a function to transform a number of input values to a range of value to between 0 ~ 1.
➔ Multinomial logistic

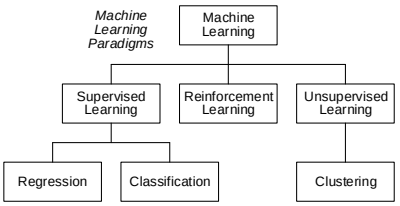
Clustering

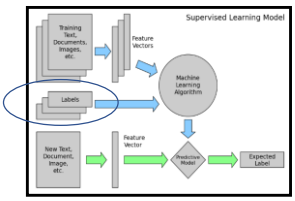
- How many groups can you divide?
  - Circle and rectangle
  - Red, yellow, blue
  - Edible and not
- This process is called ‘clustering’

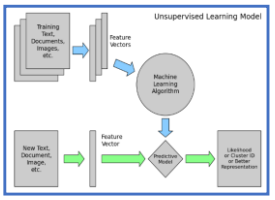
Object classification and detection



Types of learning of machine learning





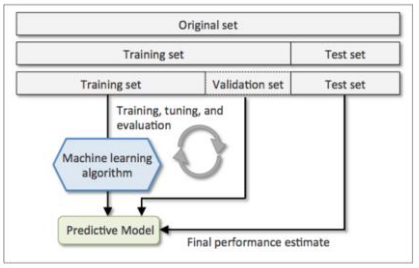


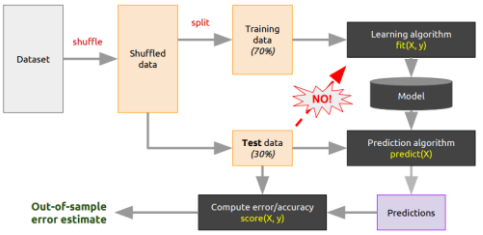
- Supervised learning (지도학습)
  - Labelled data (metrics) is already given to the computer.
  - Solving two types of problems
    - Regression problem: target variable is continuous
    - Classification problem: target variable is categorical
- Unsupervised learning (자율학습/비지도학습)
  - Finding hidden structures in datasets without any labels
  - Data is clustered using several clustering algorithms
  - Ex) Google News, Social Network Analysis, translation
- Reinforcement learning (강화학습)
  - No data given. Agent interacts with the environment calculating cost of actions.
  - Network is only provided with a grade, or score, which indicates network performance.
  - Gives reward instead of label
  - Action selection, policy learning, gaming
  - Ex) Google AlphGo

http://www.astroml.org/sklearn_tutorial/general_concepts.html

Data sets

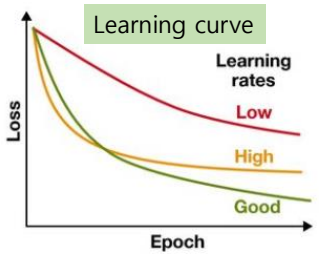
- Training set
  - data for model building by training
  - a set of examples used for learning, where the target value is known.
- Validation set
  - data for estimating error while training
  - It should not be the same as training set and used as training.
  - a set of examples used to tune the architecture of a classifier and estimate the error.
- Test set
  - data for estimating error
  - used only to assess the performances of a classifier. It is never used during the training process so that the error on the test set provides an unbiased estimate of the generalization error.

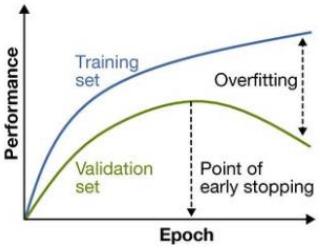




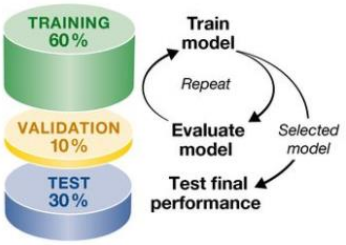
- Training error: error by training data set
- Generalization error (test error, out-of-sample error): error by test data set in order to evaluate the training model.

Training strategy

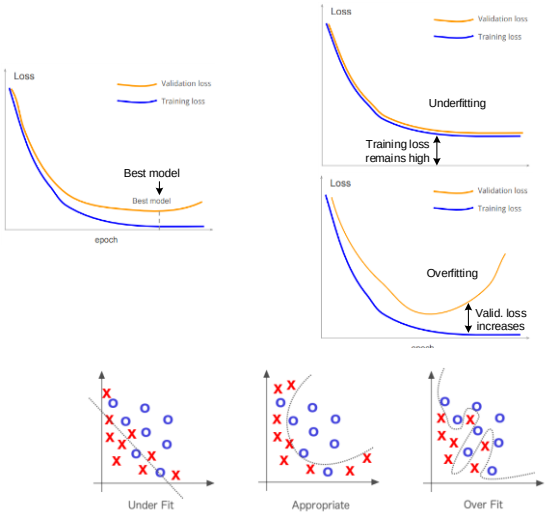




- Split data into train, validation, and test sets
  Keep 10-30% of data for validation
  - Plot learning curves as training progresses
  - Stop when validation loss starts to increase
  - Use model with minimum validation loss



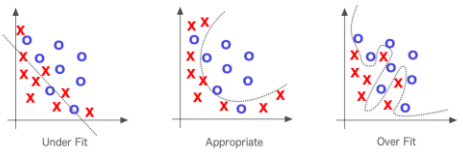
Training strategy



- Underfitting
  - Symptoms:
    - Training loss decreases at first but then stops
    - Training loss still high
    - Training loss tracks validation loss
  - Try:
    - Increase model capacity
    - Add more layers, increase layer size
    - Use more suitable network architecture
    - Use better optimization (algorithm to reduce error)
- Overfitting
  - Symptoms:
    - Training loss continues to go down
    - Validation loss decreases at first, then starts increasing
  - Try:
    - Find more training data
    - Add stronger regularization
    - dropout, drop-connect
    - Data augmentation (flips, rotations, noise)
    - Reduce complexity of the model
  - Learn the “data”, but not the underlying function

Overfitting (과적합)

- 학습 데이터 셋을 지나치게 정확하게 구분하도록 학습된 경우, 학습 데이터가 아닌 것에 대해 제대로 된 결과를 내지 못하는 문제
  - 일반화(generalization) 성능이 오히려 낮게 되는 문제가 발생

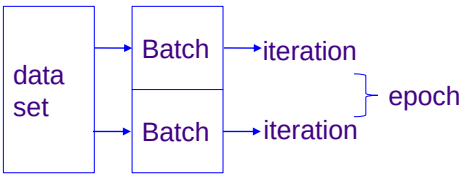


- Over-fit의 원인
  - 학습 데이터가 너무 작아서 쉽게 학습이 된 경우
  - 학습 데이터 셋이 전체 데이터 셋의 특성/분포를 반영하지 못하는 경우
  - 모델이 너무 복잡하여 대체로 많은 특징을 모델에서 반영하게 된 경우
- Parameter regularization (정규화, 일반화, 규제화, 정칙화)
  - 같은 현상을 설명하는 여러 개의 주장이 있다면 간단한 쪽을 선택한다.
  - 특정 가중치가 과도하게 반영되지 않도록 하여 모델의 일반화를 유도하는 방법
  - Regularization, Dropout, Early Stopping

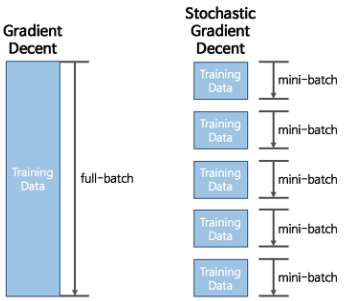
Data set, batch, epoch and iterations

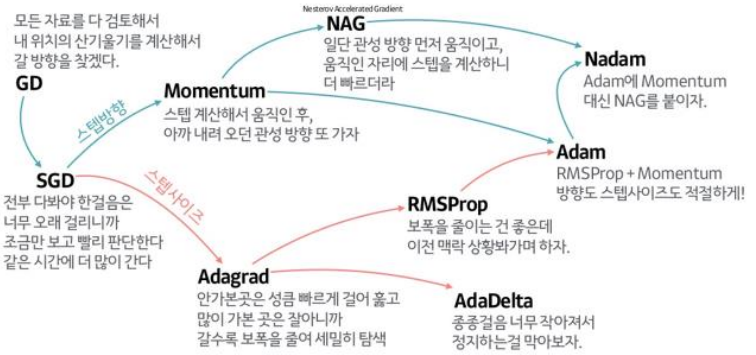
- Data set
  - training data and validation data
  - it will be large amount of data.
- Batch
  - Training data set is divided into a number of parts.
  - The part of data set is ‘batch’.
- Epoch
  - One Epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE.
  - Since, one epoch is too big to feed to the computer at once we divide it in several smaller batches.
- Iteration
  - The number of batches needed to complete one epoch

- We can divide the dataset of 2000 examples into batches of 500 then it will take 4 iterations to complete 1 epoch.
- If you have 1000 training examples, and your batch size is 500, then it will take 2 iterations to complete 1 epoch.



Optimization



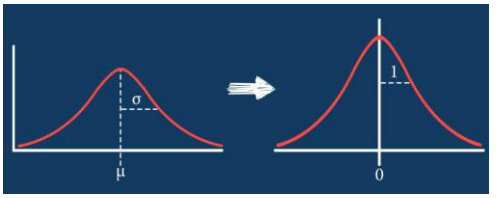


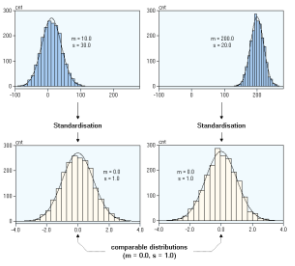
ref: 하용호: https://www.slideshare.net/yongho/ss-79607172

Normalization and standardization

- Normalization (일반화, 정규화)
  - rescales the values into a range of [0,1].
    - 데이터의 범위를 0과 1로 변환하여 데이터 분포를 조정하는 방법
  - Xchanged=(X−Xmin)/(Xmax−Xmin)

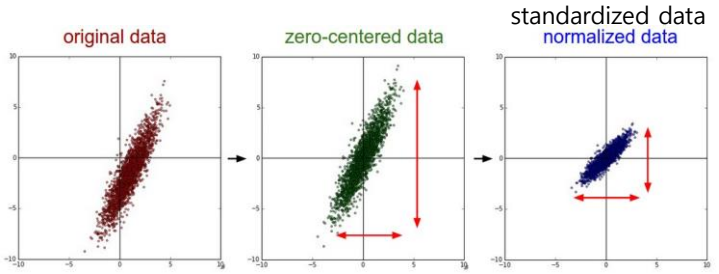
- Standardization (표준화)
  - It is also called 'normalization' in deep learning.
  - rescales data to have a mean (μ) of 0 and standard deviation (σ) of 1 (unit variance).
    - 평균을 기준으로 어느 정도 떨어져 있는지를 나타냄
  - *Xchanged=(X−μ)/σ*



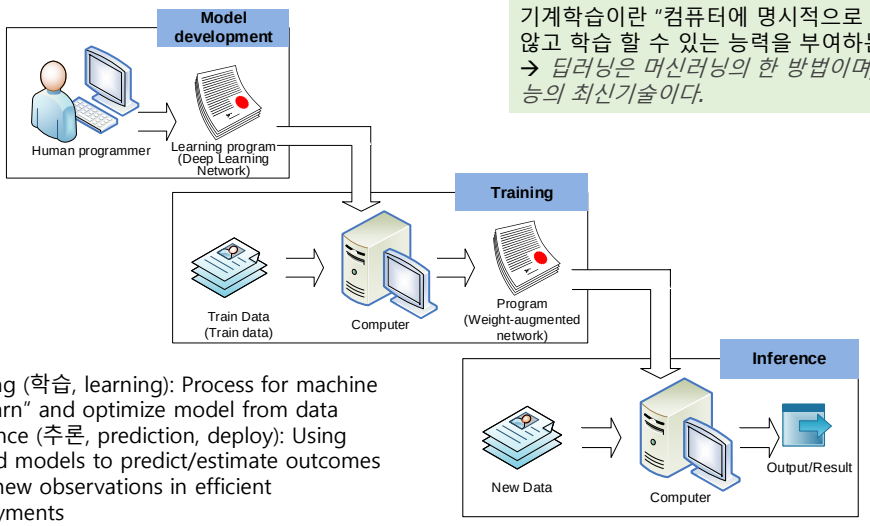


Feature scaling

- Feature scaling (data normalization)
  - making range of variables (data) comparable.
  - including normalization (re-scaling) and standardization



Deep Learning Design Flow

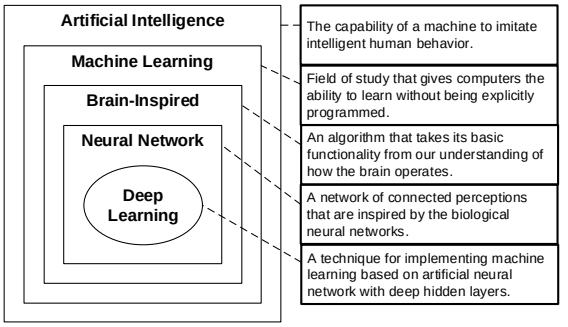


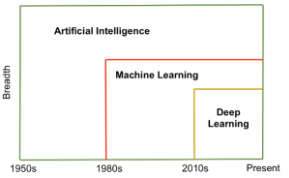
기계학습이란 "컴퓨터에 명시적으로 프로그래밍하지 않고 학습 할 수 있는 능력을 부여하는" 것
→ *딥러닝은 머신러닝의 한 방법이며, 이들은 인공지능의 최신기술이다.*

- Training (학습, learning): Process for machine to "learn" and optimize model from data
- Inference (추론, prediction, deploy): Using trained models to predict/estimate outcomes from new observations in efficient deployments

AI to DL

- AI (인공지능)
  - 지능적인 사람의 행동/판단을 모사하는 기계의 능력
- ML (기계학습)
  - 명시적인 프로그램하지 않고 컴퓨터가 배우는 능력에 대해 공부하는 영역
- BI (두뇌모사)
  - 두뇌가 작동하는 방식에 기초한 알고리즘
- NN (뉴론네트웍)
  - 생물학적 뉴론 네트웍에 기초한 퍼셉트론을 네트웍
- DL (딥러닝)
  - 많은 수의 숨은 레이어를 갖는 인공 뉴론 네트웍에 기초한 기계학습의 구현





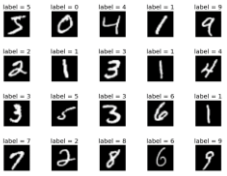
Popular Deep Neural Networks

Image classification case

	LeNet-5	AlexNet	GoogLeNet (V1)	ResNet-50	ResNet-152
Data set	MNIST	ImageNet	ImageNet	ImageNet	ImageNet
Purpose	Handwritten digit classification	Image classification	Image classificati on	Image classifica tion	Image classifica tion
Error (%)[Human]	0.95 [0.2~0.3]	16.4 [5]	6.7	5.3	3.57
Year	1998	2012	2014	2015	2015
Image size	28x28	227x227	224x224	224x224	
Layers	4	8	22	50	152
Weights	431k	61M	7M	25.5M	??
MACs	2.3M	724M	1.43G	3.9G	??
Training time		a week			
Inference time					
etc		2 GPU			

Data set: MNIST

- The MNIST database (Modified National Institute of Standards and Technology database)
  - a large database of handwritten digits that is commonly used for training various image processing systems.
- Digit classification
  - 28x28 pixels (B&W)
  - 10 classes: 0, 1, …, 9
  - training set: 60,000 training image
  - test set: 10,000 testing imag



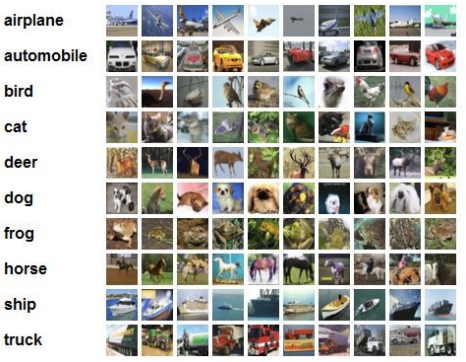
Type	Classifier	Error rate (%)
Linear classifier	Pairwise linear classifier	7.6
K-NN	K-Nearest Neighbors	0.52
Neural network	2-layer	1.6, 0.7
Deep NN (LeNet-5)	5-layer	0.95
Deep NN	6-layer	0.35
Convolutional NN	6-layer	0.31~0.21

1998: LeNet, 0.95% error
2013: ICML, 0.21% error

http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html

Data set: CIFAR-10/CIFAR-100

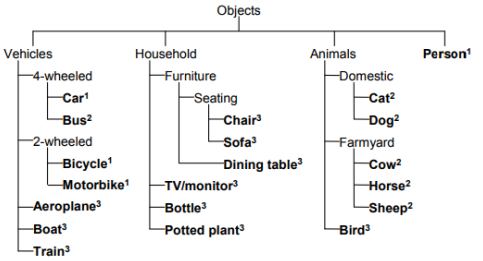
- CIFA: Canadian Institute For Advanced Research
- CIFA-10
  - Object classification
  - image dataset consists of 60,000 (32x32-pixesl/image) color images in 10 classes, with 6,000 images per class.
  - 32x32 pixels (color)
  - 10 classes containing 6,000 images each
  - 50,000 training
  - 10,000 testing
- CIFA-100
  - 100 classes containing 600 images each



https://www.cs.toronto.edu/~kriz/cifar.html

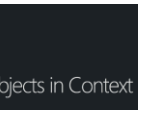
Data set: VOC2012

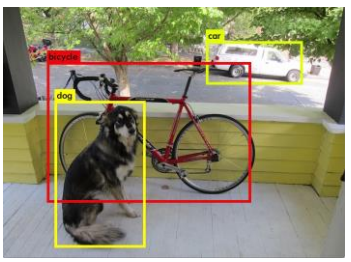
- VOC: PASCAL Visual Object Classes
  - Pattern Analysis, Statistical Modeling and Computational Learning
  - http://host.robots.ox.ac.uk/pascal/VOC/
  - VOCO: 2005~2012
  - VOC2012
    - 20 classes
    - 11k images
    - The train/val data has 11,530 images containing 27,450 ROI (region of interest) annotated objects and 6,929 segmentations.

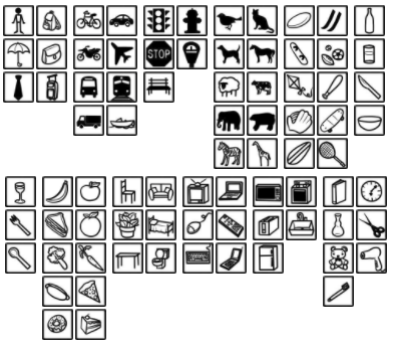


Data set: COCO

- COCO: Common Objects in Context
  - 100k images
  - 80 classes
  - detection labels
  - http://cocodataset.org/#home



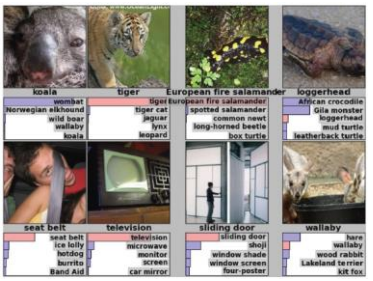




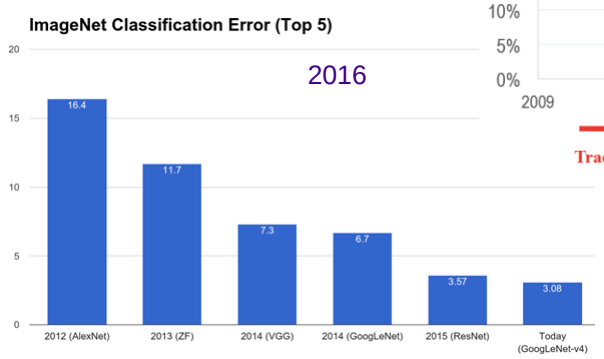
Data set: ImageNet and ILSVRC

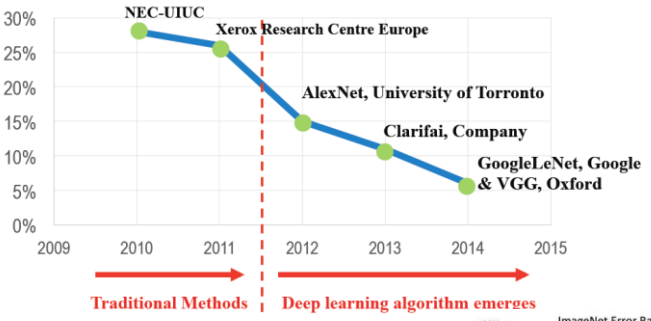
- The **ImageNet** project is a large visual database designed for use in visual object recognition software research since 2009.
  - http://www.image-net.org/
  - Over 15M labeled high resolution images
    - Annotated
  - 256x256 pixels (color)
  - Roughly 22K categories
  - Collected from web and labeled by Amazon Mechanical Turk (Mturk)

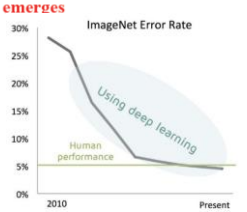
- ILSVRC: ImageNet Large-Scale Visual Recognition Challenge
  - An annual software contest run by ImageNet project since 2010
  - http://image-net.org/challenges/LSVRC/
  - 150K images, 1K object classes
  - Error by human: ~5% (classification)



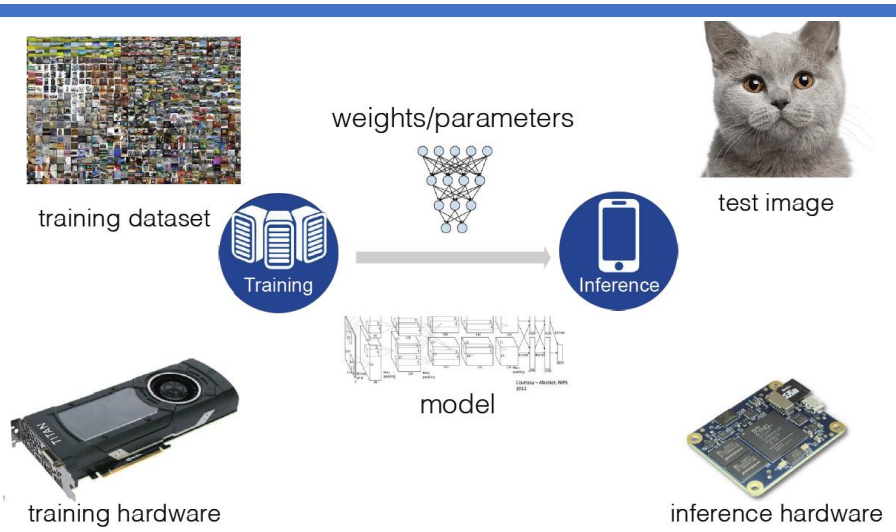
ImageNet and ILSVRC results







Training and inference



Song Han, Deep Learning Tutorial and Recent Trends, FPGA'17, 2017.

Training time

- VGG net (Karen Simonyan et. al. ICLR 2015)
  - 16/19 (CNN+FC)
  - GPU: 4 Titan Black
  - 2~3 week
- Google's Neural Machine Translation system
  - 8 layer LSTM RNN
  - GPU: 96 Tesla K80
  - ~1 week

What have been addressed

- Artificial intelligence (AI), Machine learning (ML), Deep learning (DL)
- Regression, clustering, classification
- Supervised learning, unsupervised learning
- Data set: Training set, validation set, test set, epoch, batch, iteration
- Under-fitting, over-fitting
- Model development, training, inference
- Data sets: MNIST, CIFAR-10, VOC, COCO, ImageNet