Deep Learning

Giri Iyengar

Cornell University

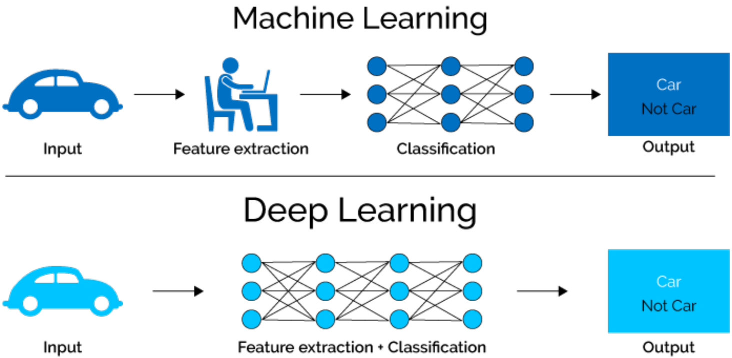
gi43@cornell.edu

Feb 12, 2018

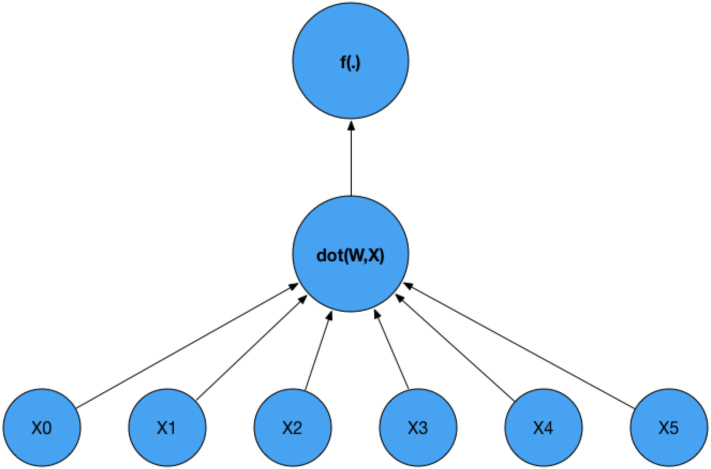
Overview

1. Deep Learning

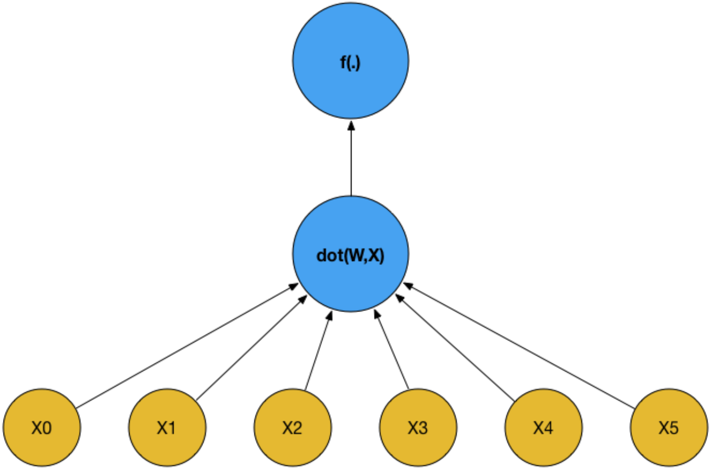
Deep Learning vs ML



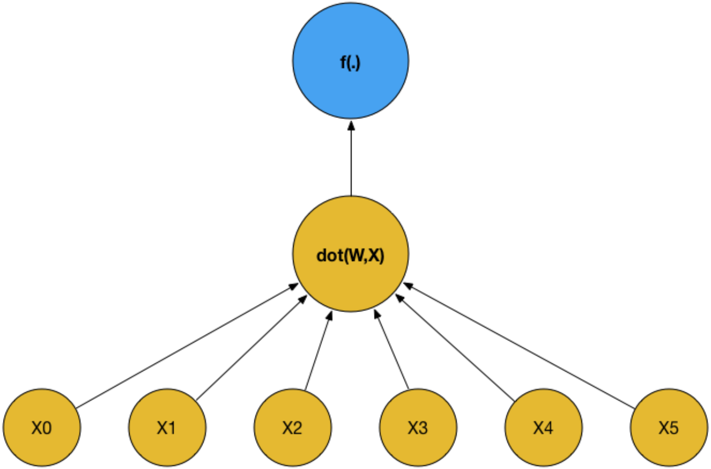
Basic Computation of the Neuron



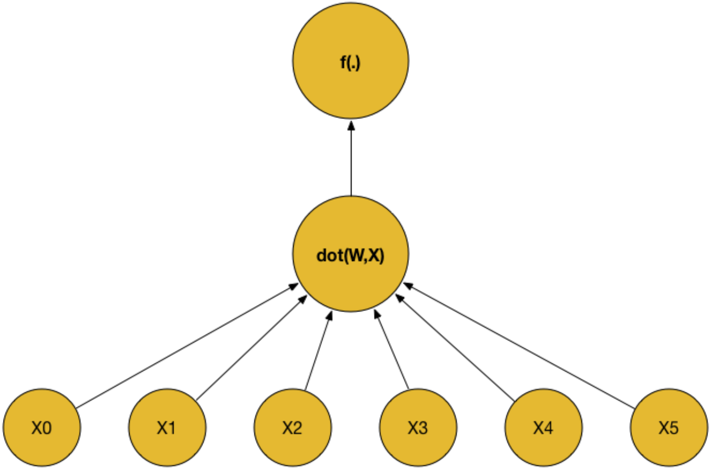
Basic Computation of the Neuron



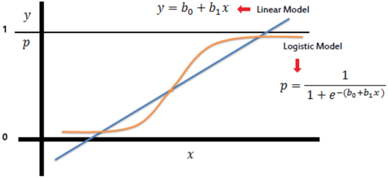
Basic Computation of the Neuron

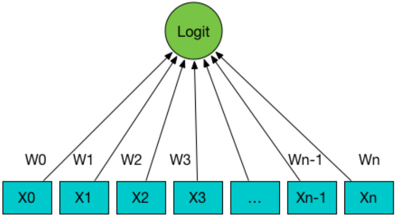


Basic Computation of the Neuron

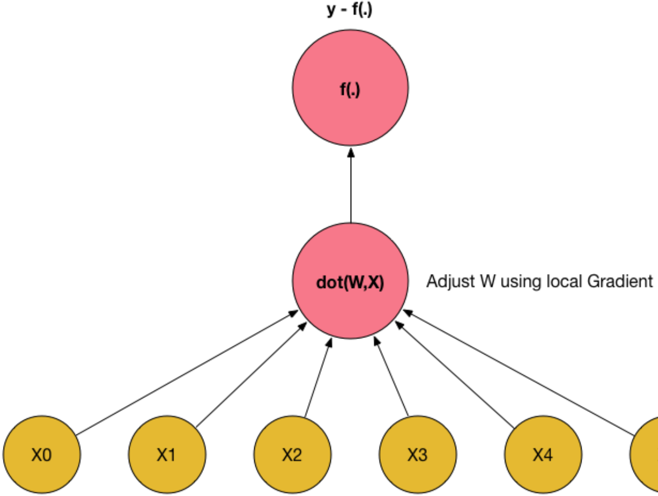


Logistic Regression is just a simple Neuron Computational Unit

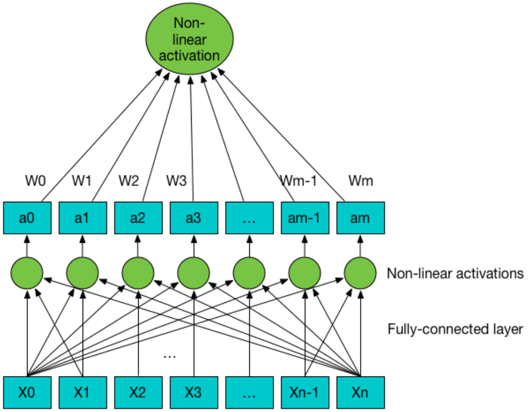




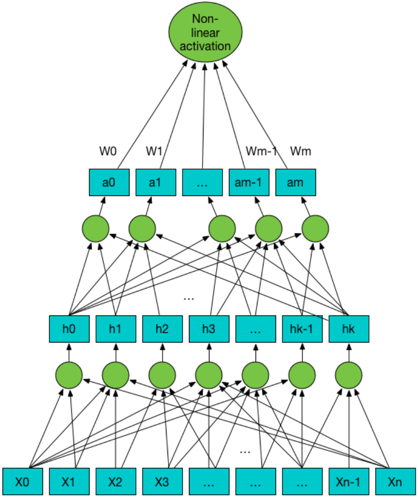
Training the Neuron



Multi Layer Perceptron



Multi Layer Perceptron



Why was this not done before?

- Too many model parameters
- Difficult to train
- Slow to train
- Prone to overfitting
- Beyond one hidden-layer, wasn't practical to fit
- Other algorithms such as SVMs had superior performance to MLP

What changed?

- Data set sizes grew exponentially
- New techniques for training were invented
  1. Stochastic Gradient Descent
  2. HogWild
  3. Stochastic L-BFGS
  4. Fast hyper-parameter search techniques
- Computational Capacity grew making training times manageable
- New algorithms together with techniques for managing model complexity (regularization) demonstrated superior performance – shot past all other models in terms of performance
- Finally delivering on **universal function approximator** promise

What Technologies make it work?

- Stochastic Gradient Descent
  - With SGD, training times of Deep Learning networks were dramatically reduced
  - Other techniques such as unsupervised initialization dramatically improved generalization ability of these networks to novel tasks
- Automatic Differentiation
  - Advances in algorithmic differentiation (not numerical differentiation or symbolic differentiation) greatly simplified writing gradient descent code for such models
- GPU
  - Availability of programmable Graphics Processing Units and compilers (CUDA) further sped up training times

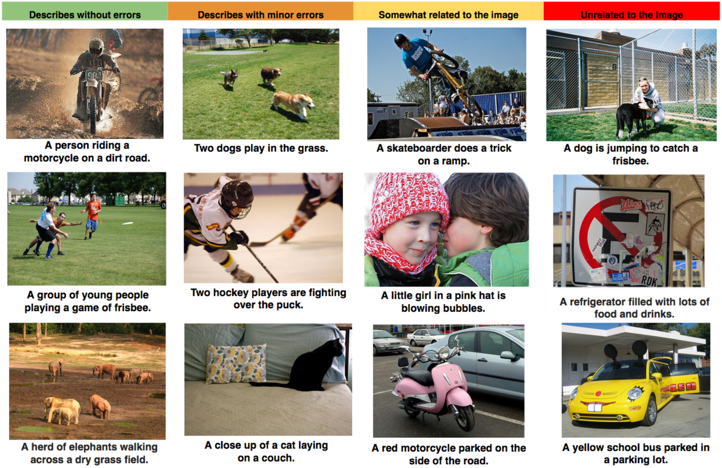
What Technologies make it work?

- Regularization
    - Advances in the understanding of regularization techniques to handle model complexity reduced overfitting issues
- Dataset sizes
    - Exponential Growth in dataset sizes further reduced overfitting issues and enabled these networks to show significantly better performance than traditional models
- Compute Capabilities
    - Improved compute capabilities make it feasible to deploy such models in production where they are able to score billions of rows a day

Some successes of Deep Learning

- Image/Object Recognition
- Speech Recognition
- Natural Language Processing
- Machine Translation

Object Recognition



Object Recognition

MS-COCO Dataset

Common Objects in Context - a new Captionining and Detection challenge

MS-COCO 2015

Some sample captions from 2015 challenge

The man at bat readies to swing at the pitch while the umpire looks on.

A large bus sitting next to a very tall building.

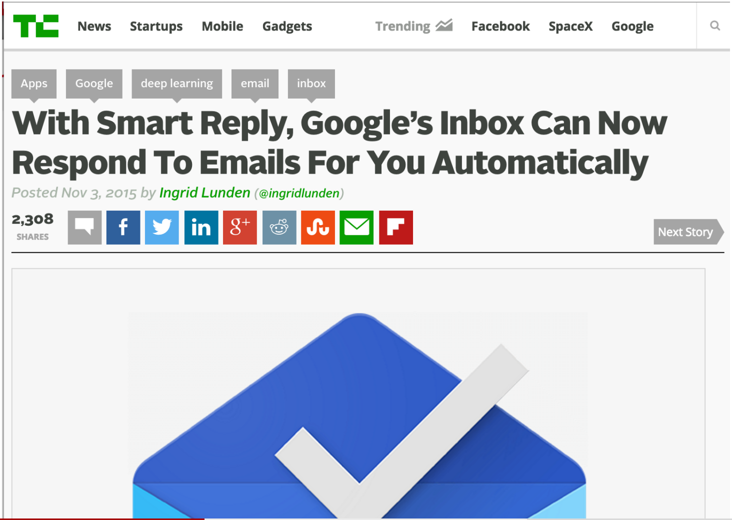
Speech Recognition, Translation, and Synthesis

- Microsoft Research Speech Breakthrough Play

Named Entity Recognition

- Tutorial on NER ▸ Go

NLP for Smart Email Replies



Some criticisms

- Theoretical Guarantees
- Models that are not well-understood


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
tapadh leat
спасибо
Баярлалаа
рахмат
bedankt
dziękuje
obrigado
sukriya
kop khun krap
go raibh maith agat
mochchakkeram
arigatō
dakujem
мерси
merci
grazie
terima kasih
감사합니다
ありがとう
dank je
hvala
mauruuru
sagolun
takk
taiku
хвала
asante
manana
enkosi
blagodaram
faafetai lava
ευχαριστώ