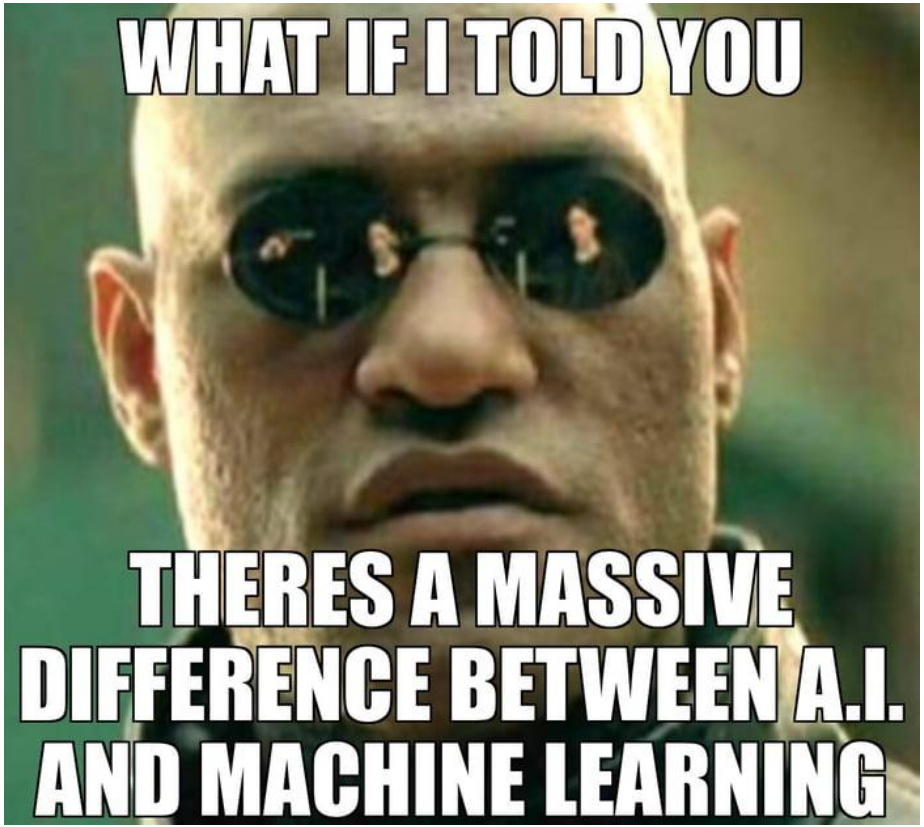
quantiphi

Intro to AI/DL/ML

Machine Learning

Machine Learning in Day to Day Life?

AI and Machine Learning



SPAM Classifier

A computer program is said to learn from **experience E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in T as measured by P, improves with experience

- **E**: I classify some of my mails into ‘spam’ or not
- **T**: The classifier puts incoming mails in ‘spam’ or not
- **P**: Ratio of the mails correctly classified automatically

A Machine Learning Approach Works Only If

1. **Some Pattern Exist in Data**

A Machine Learning Approach Works Only If

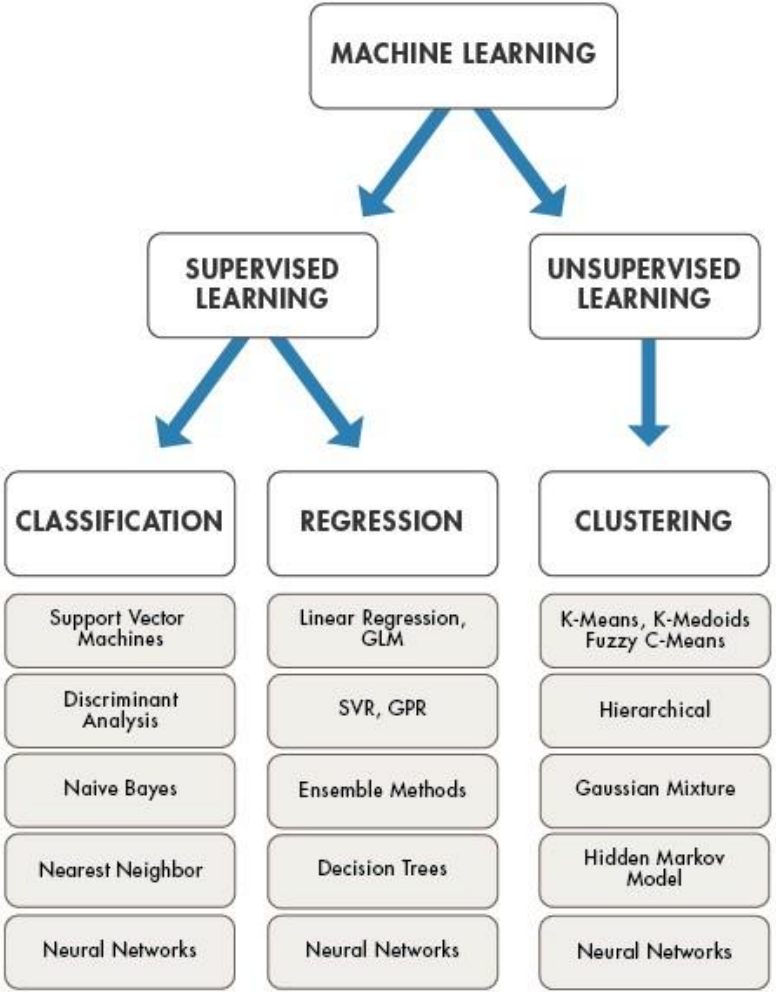
2. It's A Probabilistic Problem

A Machine Learning Approach Works Only If

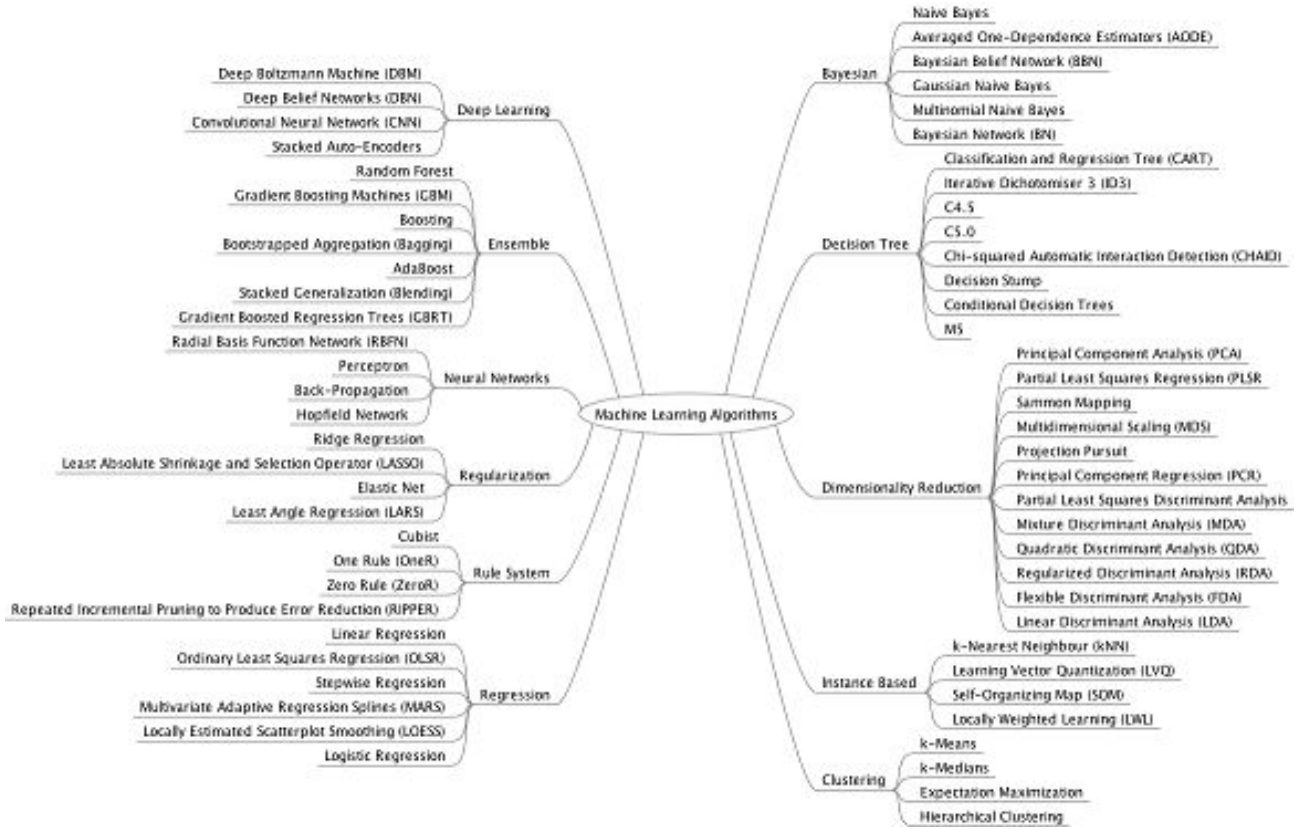
3. You have a lot of Data

quantiphi

Generalization



Algorithms


Machine Learning Algorithms
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR)
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering

Comparison

SALES FORECASTING

Develop a machine learning based forecast model to predict the sales for the upcoming years

Problem Context

- Client is a **US based pet insurance company**, providing insurance for pet owners to cover veterinary bills
- Client has been facing problems in **estimating the sales of their insurance policies**
- The goal is to assist them with a mechanism to estimate new business policy sales for 2018 using forecasting models
- Quantiphi developed a **time series forecasting model leveraging Prophet package** to forecast the sales. Seasonality based patterns were analyzed for historical data and embedded inside the model to see variance

Challenges

- ➔ Highly **inconsistent trend** in the data
- ➔ Uplift in sales due to promotions is not captured
- ➔ Customer had budgetary forecast for half of the year which was modeled based into dashboard based on their need.

Tools

Methods

Business Impact

Built forecast model to estimate the sales for 2018 which enabled the client to make better business decisions

Back to Case Study List

IDENTIFY HIGH RISK CUSTOMER GROUPS

Client wants to identify customer groups that are highly likely to churn and potential reasons behind the churn

Problem Context

- Client, an American insurance company, is the **largest provider of supplemental insurance** in the US providing financial protection to more than 50 million people worldwide.
- Client wants to identify the customer groups having higher chance to churn and better understand the reasons behind this **higher risk of churn** and levers they can pull to **mitigate the risk of defection**
- Quantiphi created a model that **outperformed client's existing model** in identifying the customer groups that are at a high risk of churning out and also suggested interpretable reasons behind the same

Challenges

- ➔ Only 2 years data made the train and test split difficult
- ➔ Very few independent metrics available to capture trends or take a top down approach
- ➔ Dataset available was highly imbalanced

Tools

Methods

Business Impact

Enable developing an action list for the business to prevent loss of customer groups and potential acceleration of business growth, by providing intuitive recommendations

Back to Case Study List

IDENTIFY HIGH RISK CUSTOMER GROUPS

Client wants to identify customer groups that are highly likely to churn and potential reasons behind the churn

Problem Context

- Client, an American insurance company, is the **largest provider of supplemental insurance** in the US providing financial protection to more than 50 million people worldwide.
- Client wants to identify the customer groups having higher chance to churn and better understand the reasons behind this **higher risk of churn** and levers they can pull to **mitigate the risk of defection**
- Quantiphi created a model that **outperformed client's existing model** in identifying the customer groups that are at a high risk of churning out and also suggested interpretable reasons behind the same

Challenges

- ➔ Only 2 years data made the train and test split difficult
- ➔ Very few independent metrics available to capture trends or take a top down approach
- ➔ Dataset available was highly imbalanced

Tools

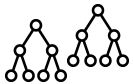
RStudio

Apache Zeppelin

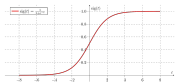
RDS

EC2

Methods



Randomforest



Logistic Regression

dmlc XGBoost

Business Impact

Enable developing an action list for the business to prevent loss of customer groups and potential acceleration of business growth, by providing intuitive recommendations

2x

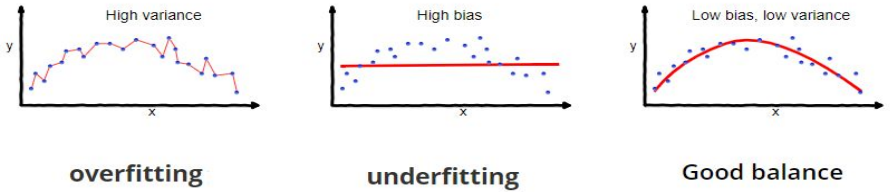
2.7x Improvement over client's current, leveraged model

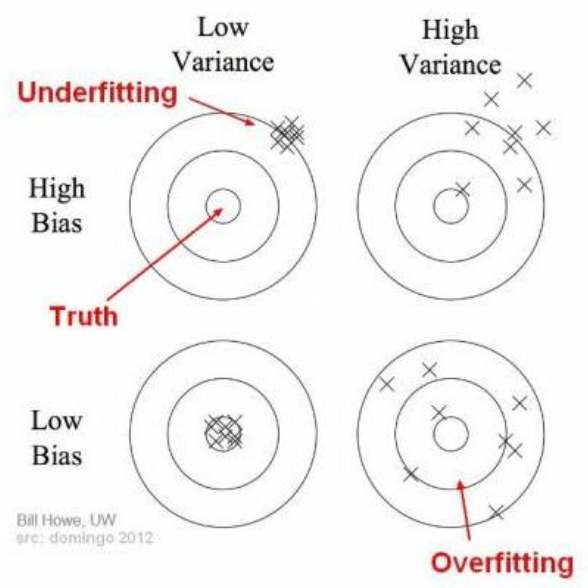
Back to Case Study List

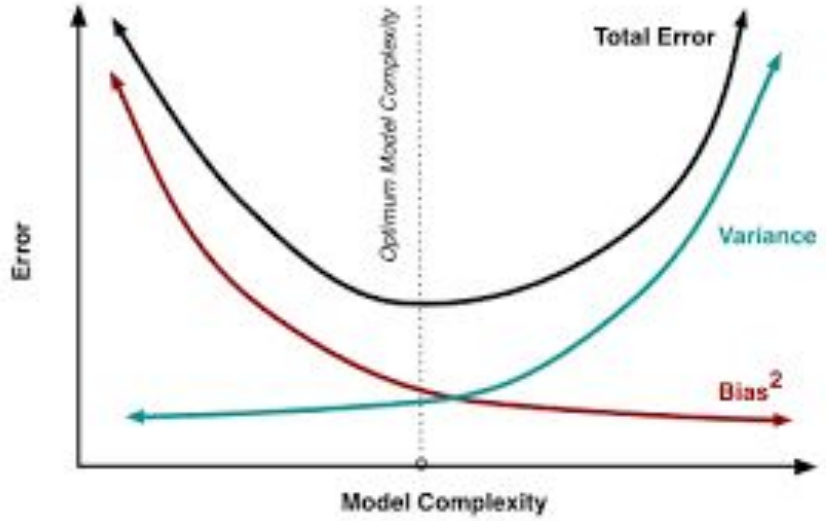
Bias Variance Trade Off

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

$$Err(x) = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$







A visual introduction to machine learning

http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Key Terms

Bootstraping

Bagging

Confusion Matrix

Overfitting

Ensemble

Imbalanced Data

Train- Test-Validation

Cross Validation

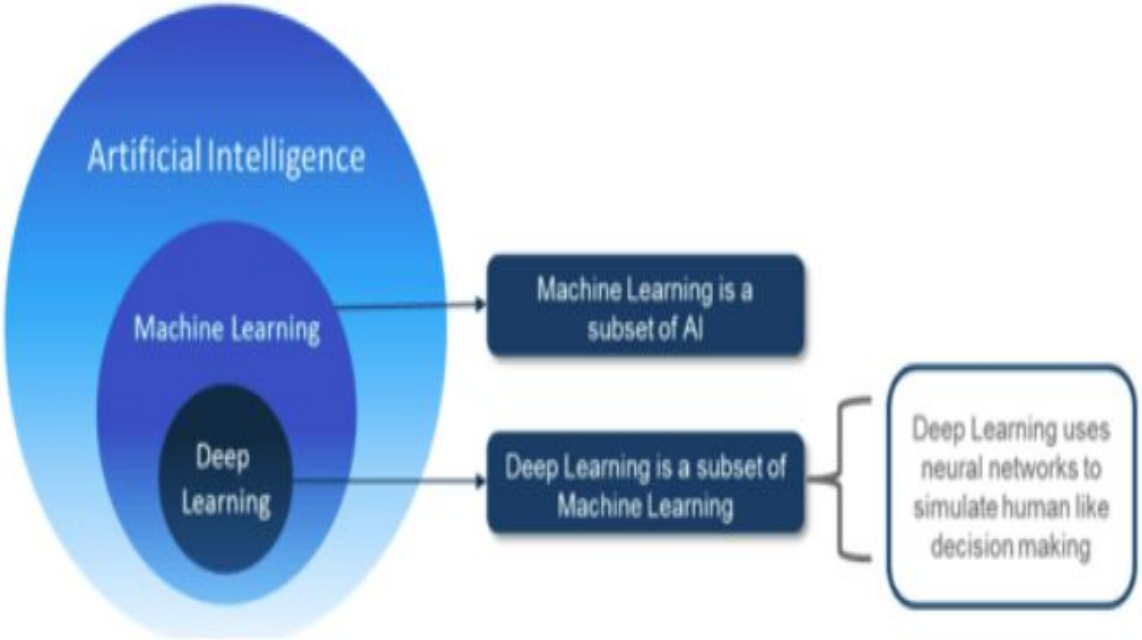
RMSE

Deep Learning

Deep Learning is at the cutting edge of what machines can do…

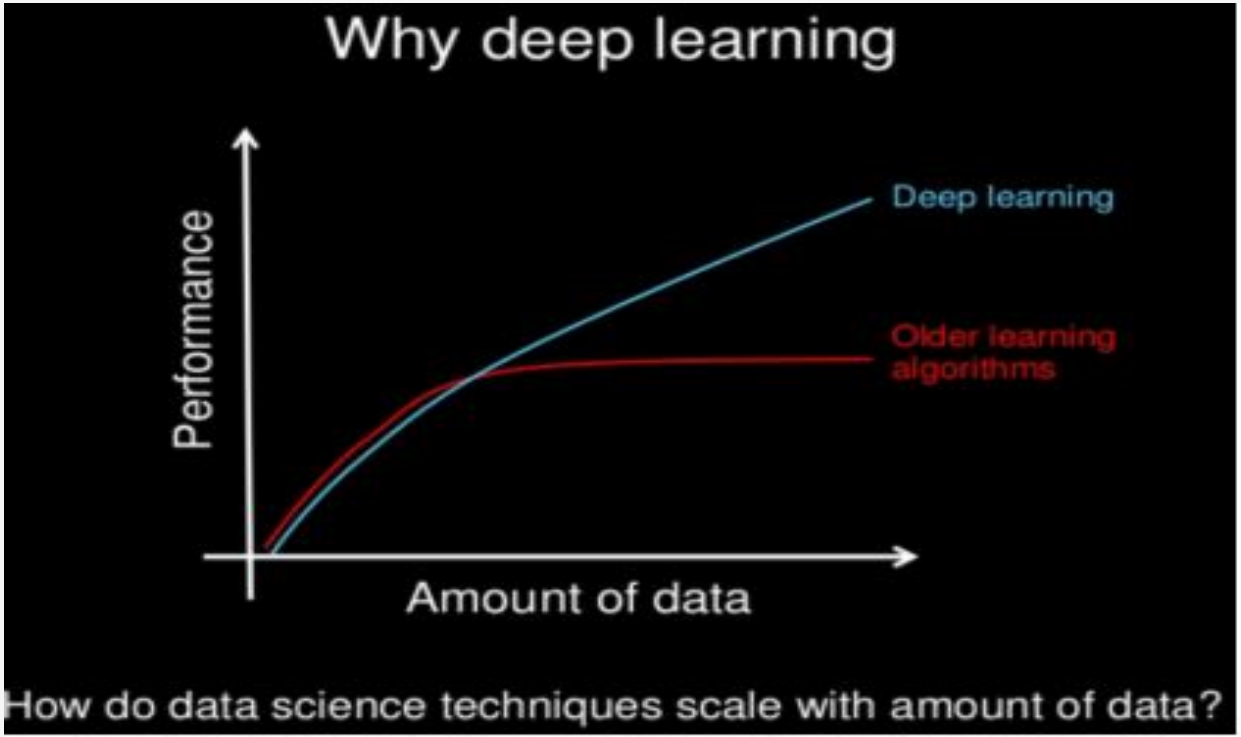
What is Deep Learning?

- Fasted growing segment of AI
- Uses learning algorithms that try to mimic the human brain
- Requires massive amounts of data



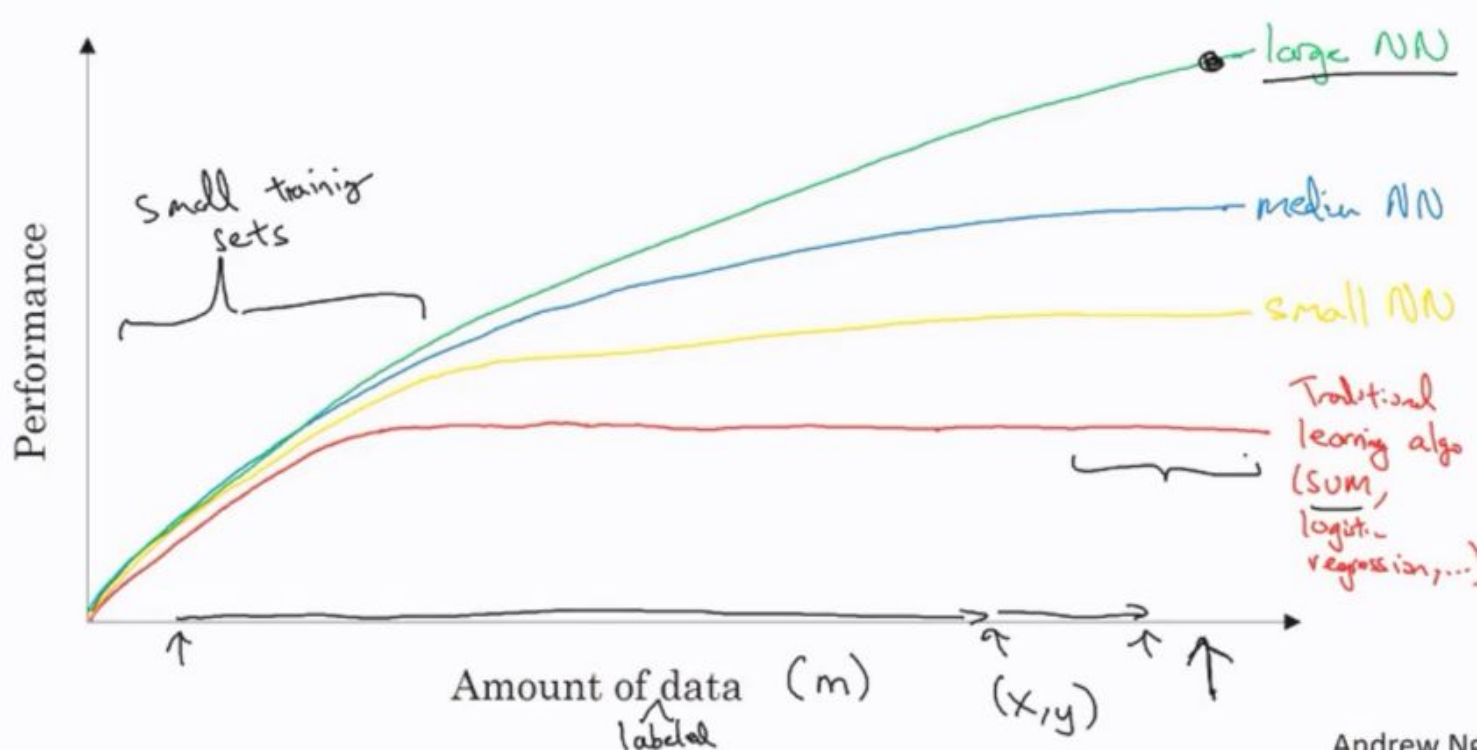
quantiphi

Why is Deep Learning important?

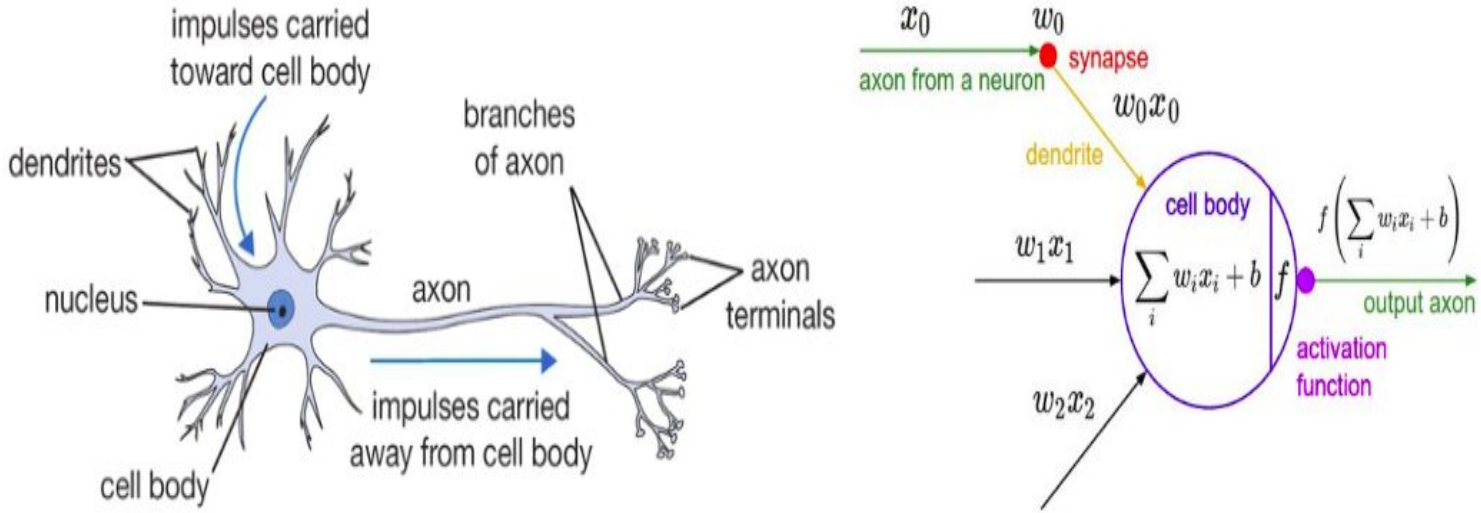


Why the recent buzz?

Scale drives the growth



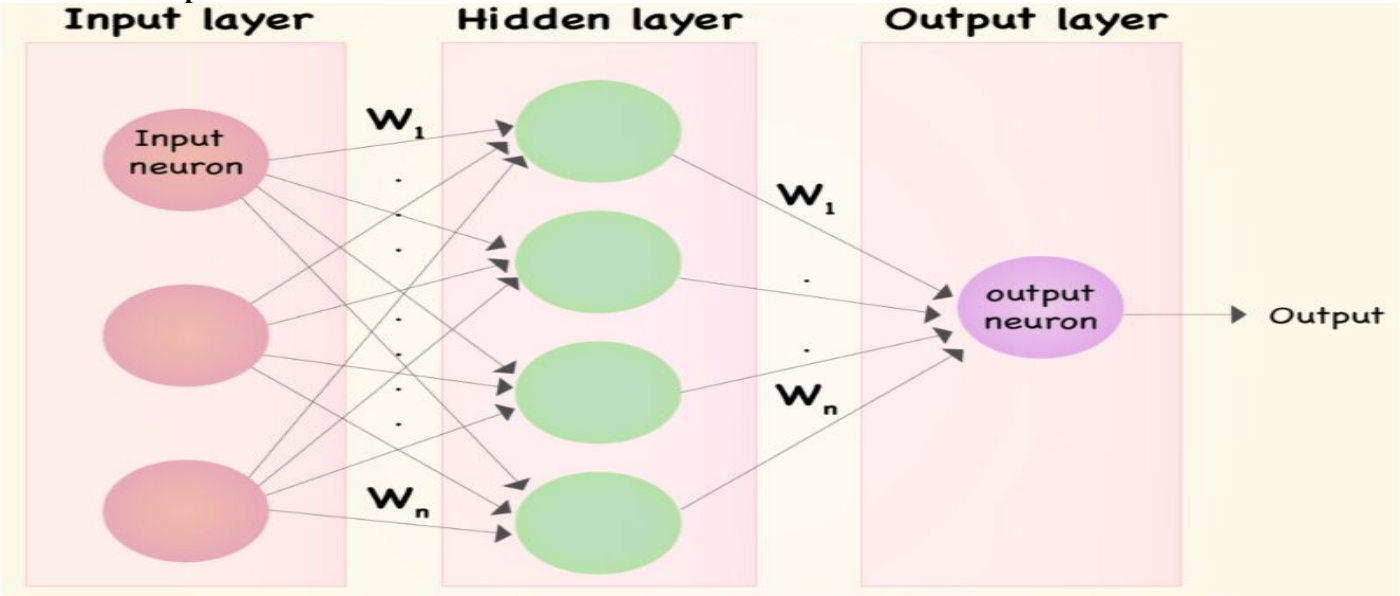
Biological motivation and connections



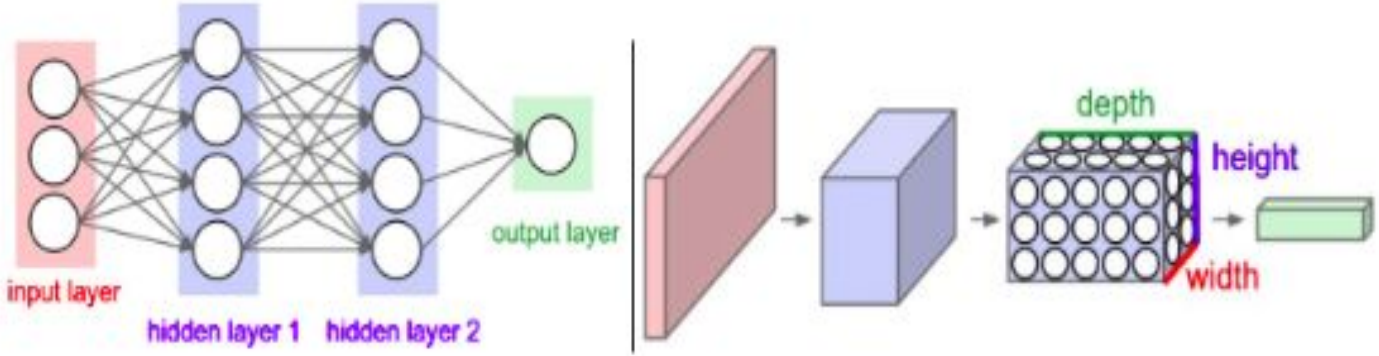
biological neuron (left) and a common mathematical model (right)

What is a Neural Network?

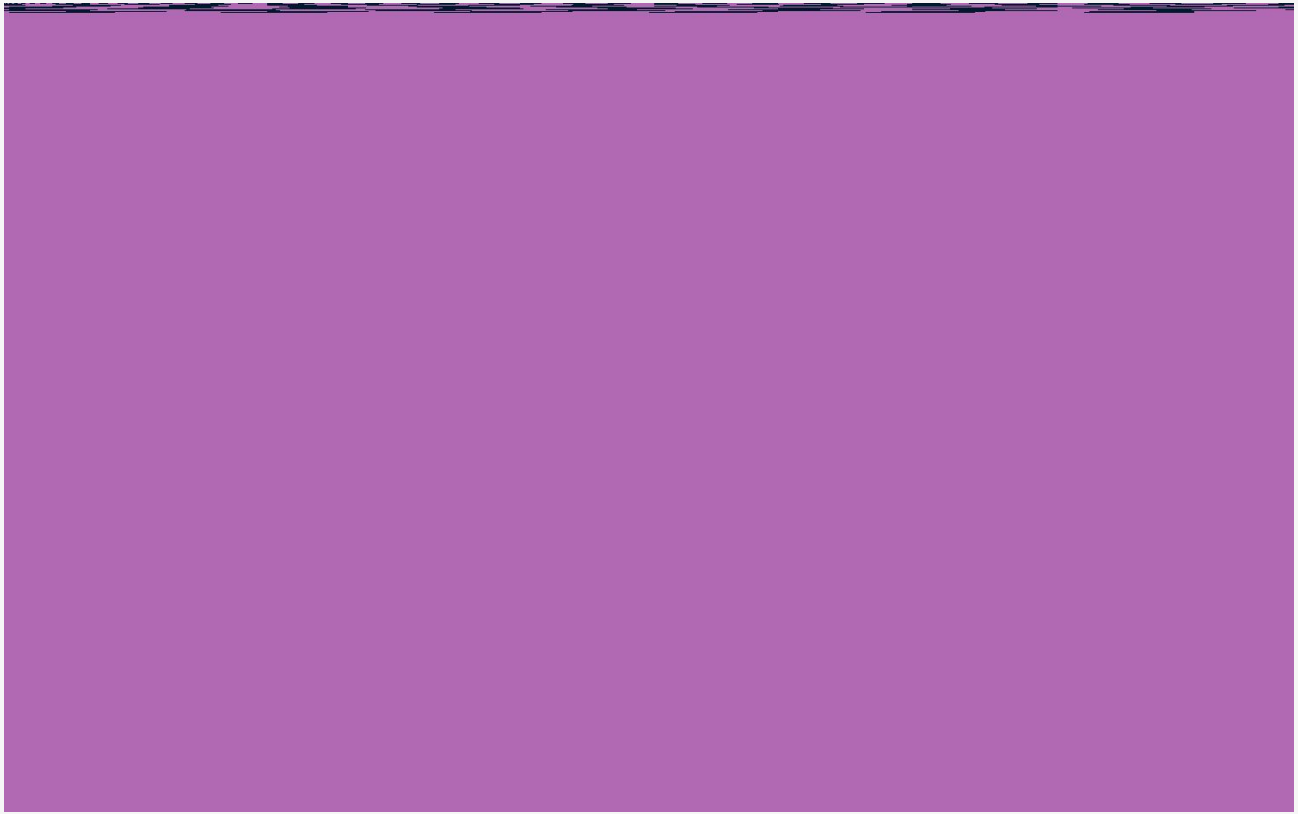
- Computing system with simple but highly interconnected nodes called **Neurons.**
- Processes information in a similar way the human brain does
- They have the ability to learn and model these non-linear and complex relationships.



Standard NN Vs CNN



Building the intuition

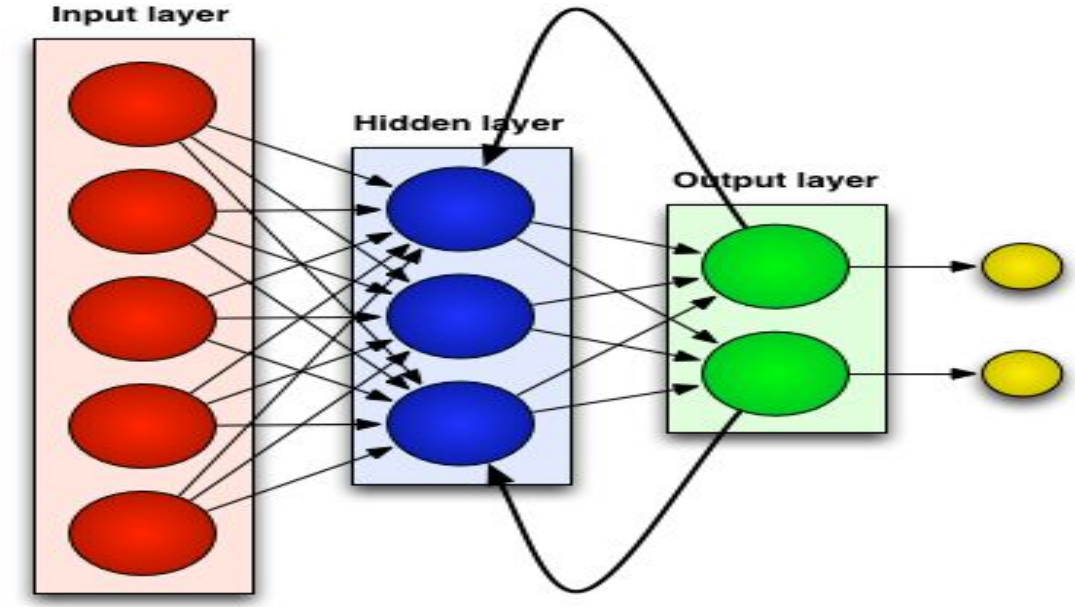


Convolutional Neural Network

- **CNNs have two components**:
  - **The Hidden layers/Feature extraction part** - Performs a series of convolutions and pooling operations during which the features are detected.
  - **The Classification part** - The Fully connected layers serve as a classifier on top of these extracted features and assigns a probability for the predicted objects.

Recurrent Neural Networks

- State of the art algorithm for sequential data and the only ones with an internal memory.
- When it makes a decision, it takes into consideration the current input and also what it has learned from the inputs it received previously.



Applications

Speech Recognition

Digital Assistants

Natural Language Processing

Image Colorization

Health Diagnostics

Self Driving Cars

Most Popular Frameworks

PYTORCH

Keras

theano


Thank you!

About us

Quantiphi is a category defining Data Science and Machine Learning software and services company focused on helping organizations translate the big promise of Big Data and Machine Learning technologies into quantifiable business impact.

ba-training@quantiphi.com