Introduction of Machine Learning

Kevin Lee at klee@clindatainsight.com, Peng Yang at pyang@clindatainsight.com

Definition of Machine Learning

An application that provides systems (machines) the ability to automatically learn and improve from experience without being explicitly programmed.

Difference between Explicit Programming and Machine Learning

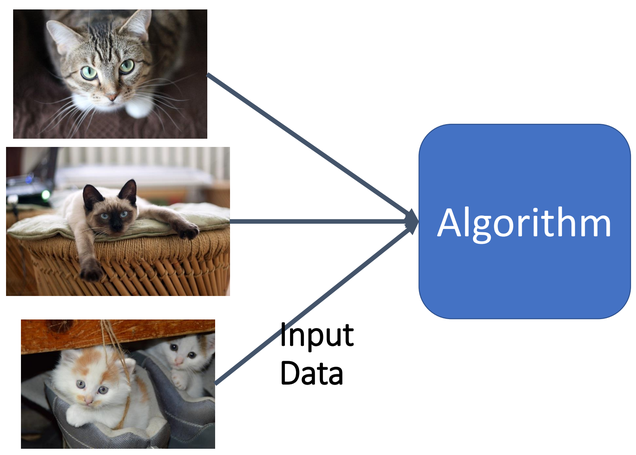
Explicit Programming

If AGE < 18, AGEGR1 = "<18";
If AGE ge 18, AGEGR1 = ">=18";

Machine Learning

AGE = 17 & AGEGR1 = "<18",
AGE = 16 & AGEGR1 = "<18",
AGE = 18 & AGEGR1 = ">=18",
,,,,,,,
With many data, AGEGR1 ="<18" or ">=18"

How does Machine Learn?



Why ML so popular?

- Cost Effective
- Complementing current services
- Customized Solution
- 24/7
- Accurate
- Solving complex problems

Machine Learning or AI Implementation

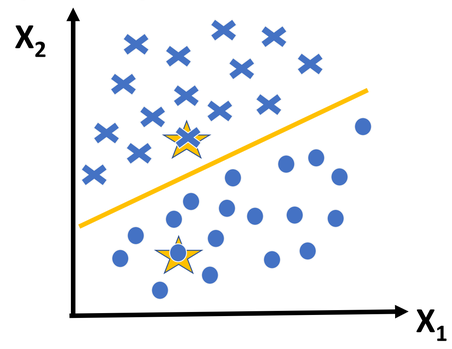
- Current Implementation
  - Voice Recognition – Siri, Alexa, Google Home
  - Recommendation – Amazon, Netflix, Spotify
  - Customer Service – Chatbot
  - Autonomous vehicle
  - Fraud Detection
  - Security Surveillance
  - Image Recognition
  - Purchase Prediction
- Current Application
  - In 2016, companies invested around $40B in AI/ML.
  - Adopters – 20%, early adopters and experimenting – 40%, thinking – 40%
- Future Impact
  - By 2030, according to McKinsey, **as many as 375 million workers (14 % of global workforce)** may need to switch jobs due to the automation by ML/AI.
  - Andrew Ng's – **"Pretty much everything that a normal person can do in < 1 sec can be automated."**

Healthcare Market for Machine Learning / AI

- US and Europe – $600M in 2016
- 40% annual rate
- $10B by 2024

Machine Learning Type

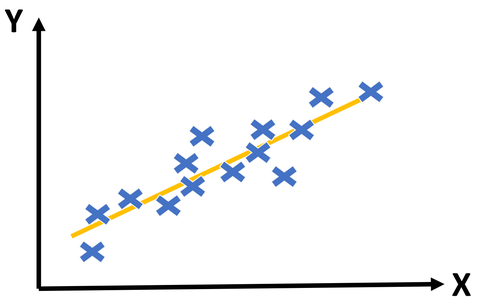
- **Supervised Machine Learning**
  - Input data labeled – has correct answers
  - Specific purpose
  - Types
    - Classification
      - Categorical Target
      - Yes/No, Mild/Moderate/Severe, 0 to 9
      - Logistic Regression, SVM, Decision Tree



- Sample Codes in Python

```
#prepare train and test datasets
x_train = …
y_train = ….
x_test = ….

#import ML algorithm
from sklearn.linear_model import
LogisticRegression

#select and train model
Log_Reg = LogisticRegression()
Log_Reg.fit(x_train, y_train)

#predict output and validate
predicted = Log_Reg.predict(x_test)
```

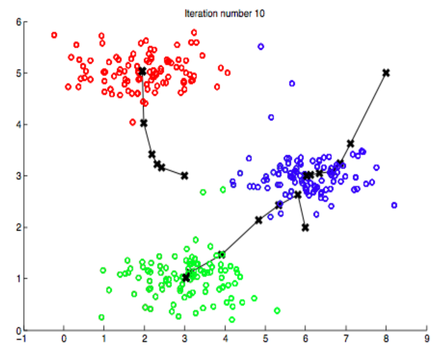
- Regression
  - Numeric Target
  - Continuous Variables
  - Predicting house price per sqft
  - Linear Regression



- Sample codes in Python

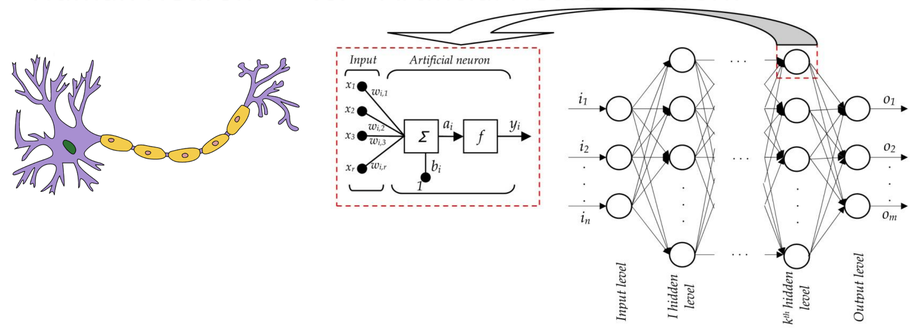
```
#prepare train and test datasets
x_train = …
y_train = ….
x_test = ….

#import ML algorithm
from sklearn import linear_model

#select and train model
linear = linear_model.LinearRegression()
linear.fit(x_train, y_train)

#predict output and validate
predicted = linear.predict(x_test)
```

- **Unsupervised Machine Learning**
  - Input data not-labeled – no correct answers
  - Exploratory
  - Clustering – the assignment of set of observations into subset (clusters)



Artificial Neural Network (ANN)

- Most powerful ML algorithm
- Game Changer
- Works very similar to human brain – neural network
- Human Neuron vs. Artificial Neural Network



- Python Codes for Deep Neural Network

```
#import ANN - TensorFlow
Import tensorflow as tf

#Set up Variables & Hidden Layers & Units
X = tf.placeholder(..)
Y = tf.placeholder(..)
hidden1 = tf.layer.dense(X, 4, activation=tf.nn.relu)
hidden2 = tf.layer.dense(hidden1, 2, activation=tf.nn.relu)
logits = neuron_layer(hidden2, 2)
….
#Set up Loss Function and Optimizer
loss = tf.reduce_mean(….)
optimizer = tf.train.GradientDescentOptimezer(0.1)
traing_op = optimizer.minimizer(loss)

#Train DNN
tf.Session.run(training_op, feed_dict={X:x_train, Y:y_train})
```

Machine Learning or AI Pharma Application

- Pharma Application
  - Drug Discovery
  - Drug Candidate Selection
  - Clinical System Optimization
  - Medical Image Recognition
  - Medical Diagnosis
  - Data Anomality
  - Personalized Medicine (Genome)
  - Sales & Marketing
- Adoption of ML/AI in Pharma
  - Slow
  - Regulatory Challenge
  - Black Box Feature
  - Cultural Challenges – Status Quo
  - Huge Upsize/ Promising
  - ML can develop SDTM, ADaM and TFL?