A Toy Model for Homomorphic Encryption for Linear Regression

Adel SOHBI

Seoul Artificial Intelligence Meetup

28th April 2018

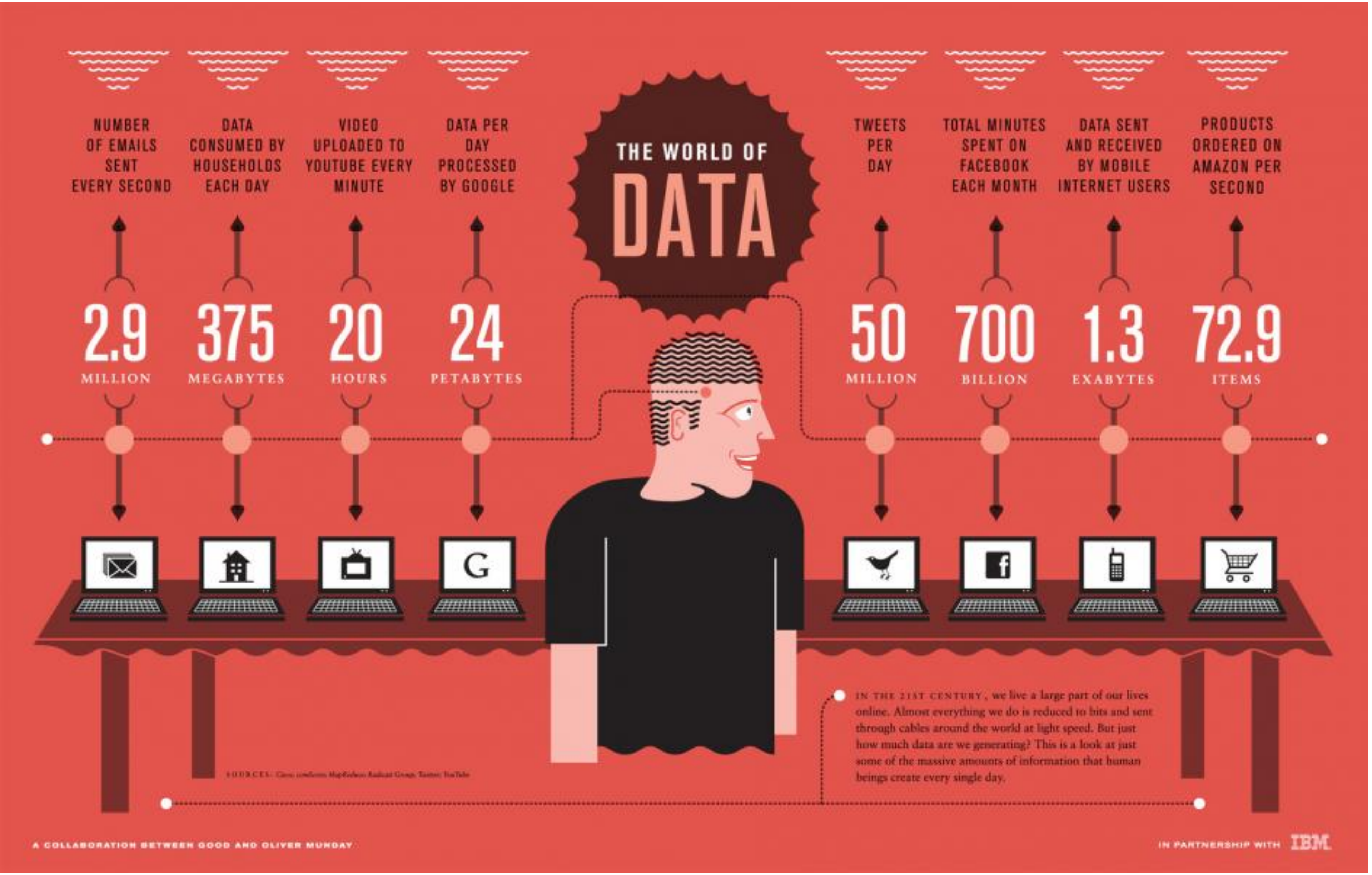
/ ! \ Disclaimer / ! \

- The model provided in this talk is a **simple toy model** to give a naive approach of Machine Learning based on Homomorphic Encryption.

- For a more **insightful** approach:

Secure Multiple Linear Regression Based on Homomorphic Encryption

Rob Hall[1], Stephen E. Fienberg[1] and Yuval Nardi[2]



THE WORLD OF
DATA
NUMBER OF EMAILS SENT EVERY SECOND
2.9
MILLION
DATA CONSUMED BY HOUSEHOLDS EACH DAY
375
MEGABYTES
VIDEO UPLOADED TO YOUTUBE EVERY MINUTE
20
HOURS
DATA PER DAY PROCESSED BY GOOGLE
24
PETABYTES
TWEETS PER DAY
50
MILLION
TOTAL MINUTES SPENT ON FACEBOOK EACH MONTH
700
BILLION
DATA SENT AND RECEIVED BY MOBILE INTERNET USERS
1.3
EXABYTES
PRODUCTS ORDERED ON AMAZON PER SECOND
72.9
ITEMS
G
IN THE 21ST CENTURY, we live a large part of our lives online. Almost everything we do is reduced to bits and sent through cables around the world at light speed. But just how much data are we generating? This is a look at just some of the massive amounts of information that human beings create every single day.
A COLLABORATION BETWEEN GOOD AND OLIVER MUNDAY
IN PARTNERSHIP WITH IBM

Cloud Wars

Q4 2017 Cloud Revenue		
IBM	$5.5B	+30%
Microsoft	$5.3B	+56%
Amazon	$5.1B	+45%
Salesforce.com	$2.68B	(qrtr ending Oct. 31)
Oracle	$1.5B	(qrtr ending Nov. 30)
SAP	$1.24B	
Google	$1B	

Source: @bobevansIT

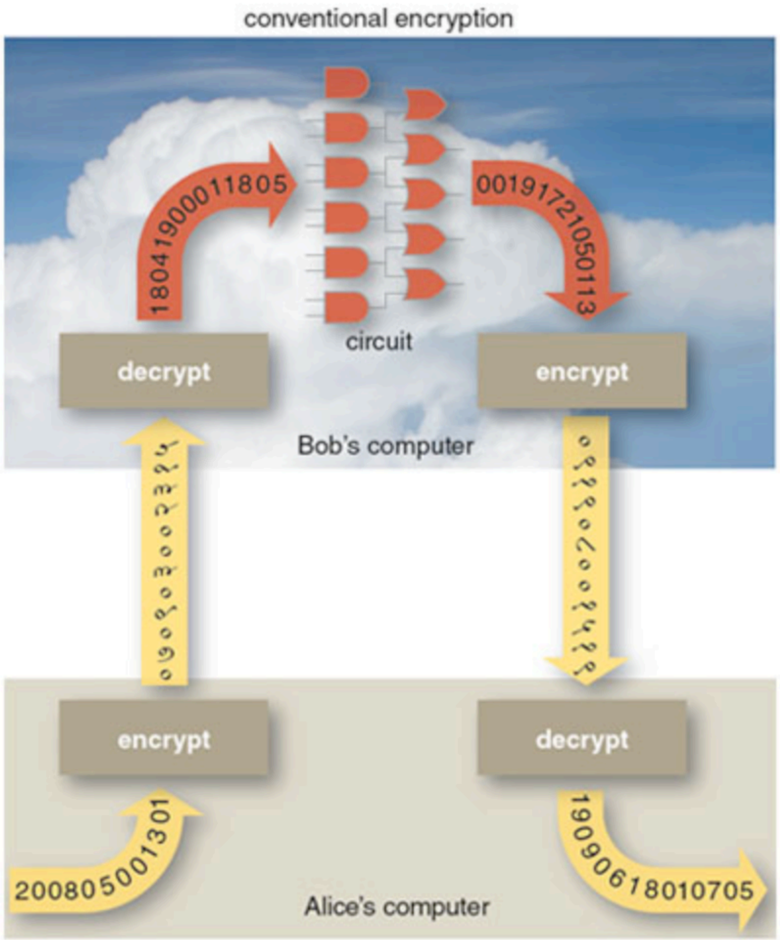
Information Privacy

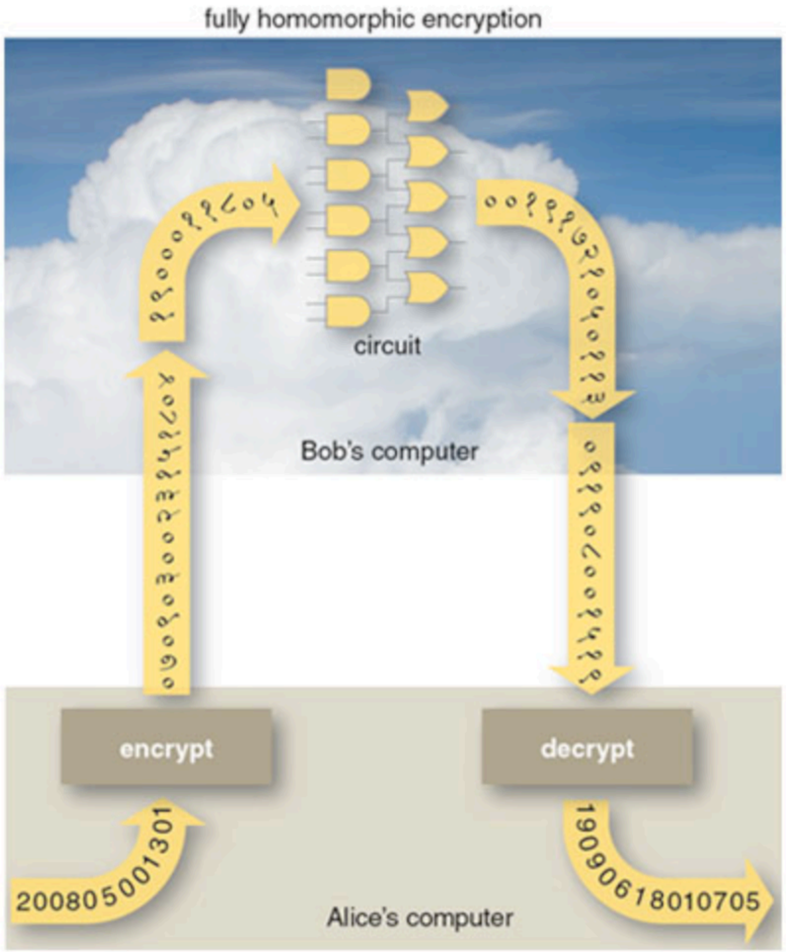
Analytics

Some self analytics about Facebook usage start appearing after a while:

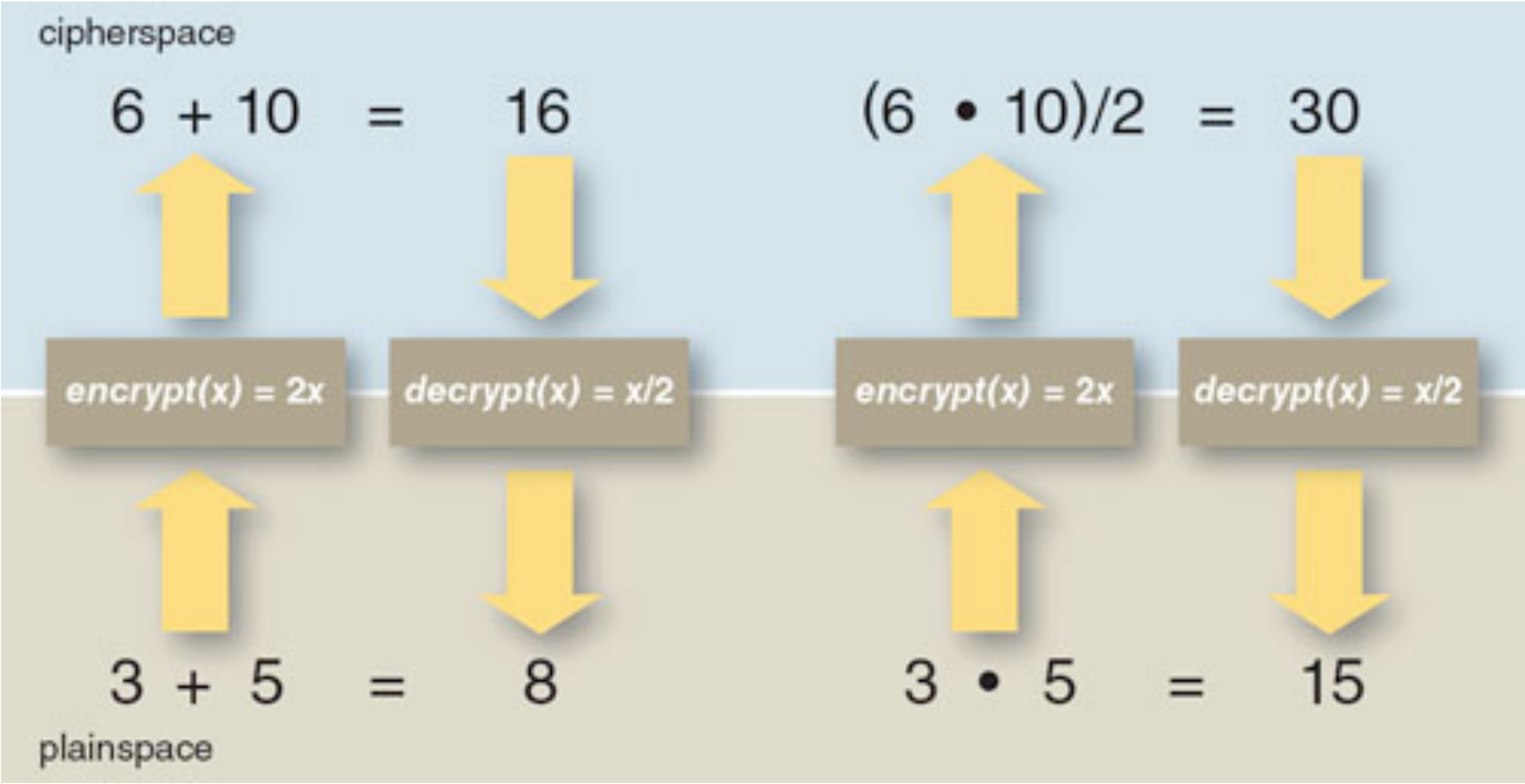
- Top friends
- Activity
- Top pages
- Top likes
- Object detection (YOLO)
- Sentiment Analysis (Watson)
- Categories
- Personality prediction
- Religious Orientation
- Political Orientation
- Other Predictions
- Shopping Preferences
- Health + Activity + Other preferences

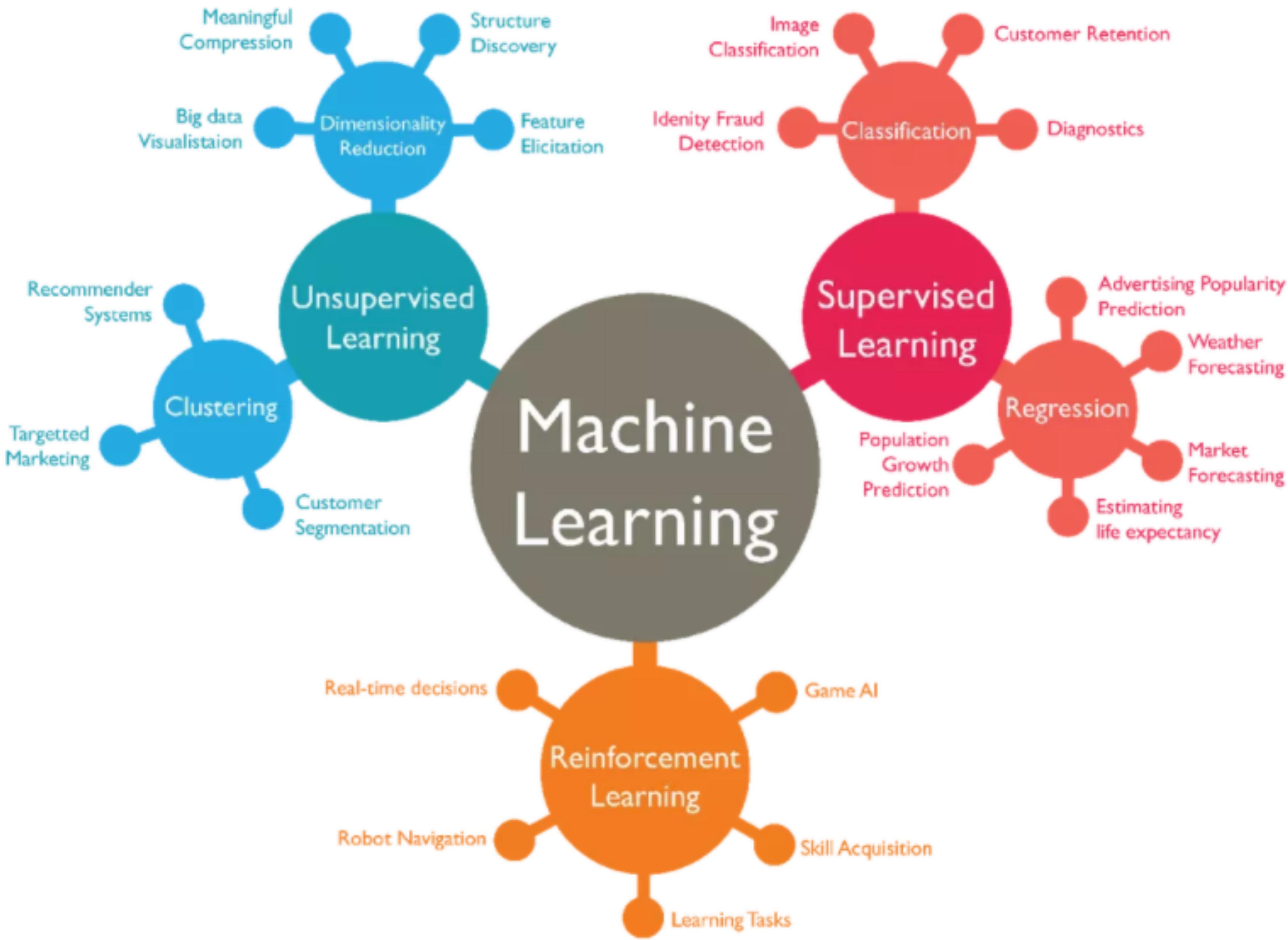
Homomorphic Encryption





Homomorphic Encryption

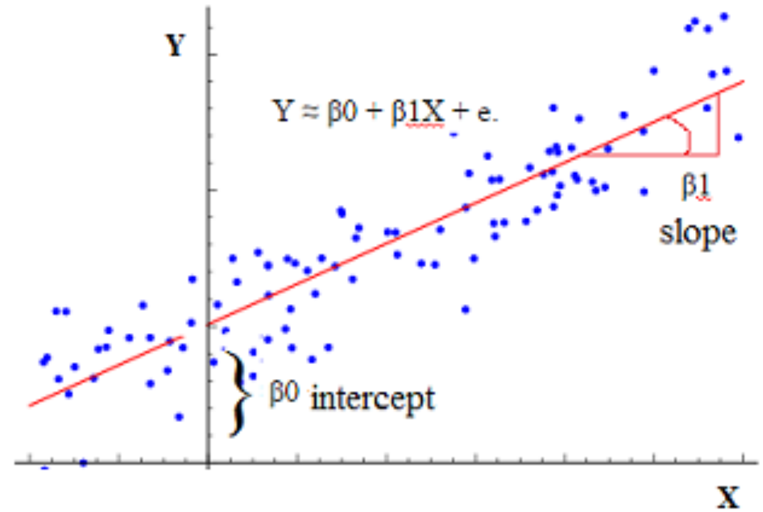



Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Linear Regression

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$Y = X\beta + \varepsilon$$



Aim of the rest of this talk

- Construct a simple **linear regression model** with the following requirements:

  1 - The data to train the model are encoded

  2 - The trained model is encoded

  3 - The model can be used with encoded data which gives an encoded result

Encoding Method

- **Symmetric-key** algorithm

- Define:
  - U_1 an semi-orthogonal matrix.
  - U_2 an invertible matrix.

- Encoded Data:

$$\widetilde{X} = U_1 X U_2$$

$$\widetilde{Y} = U_1^T Y$$

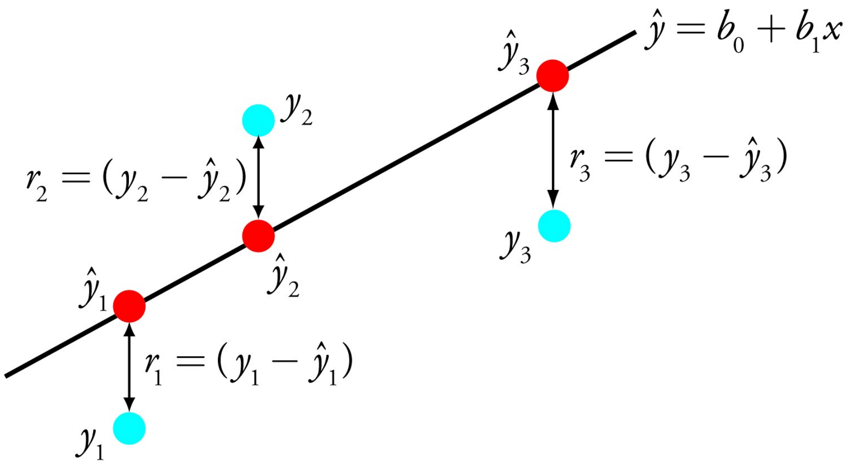
Training the Model

- **Ordinary least squares** (OLS)

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \qquad \hat{y} = X\hat{\beta}$$

- **Encoded** model:

$$\widetilde{\beta} = U_2\hat{\beta}$$



Use the Model

1 – Encode the new data with U_3 an invertible matrix:

$$\widetilde{X}_{test} = U_3 X_{test} U_2^{-1}$$

2 – Compute:

$$\widetilde{y}_{test} = \widetilde{X}_{test} \widetilde{\beta}$$

3 – Decode the result:

$$y_{test} = U_3^{-1} \widetilde{y}_{test}$$

Aim of the rest of this talk

- Construct a simple **linear regression model** with the following requirements:

  1 - The data to train the model are encoded

  2 - The trained model is encoded

  3 - The model can be used with encoded data which gives an encoded result

Aim of the rest of this talk

- Construct a simple **linear regression model** with the following requirements:

  1 - The data to train the model are encoded

  2 - The trained model is encoded

  3 - The model can be used with encoded data which gives an encoded result

Aim of the rest of this talk

- Construct a simple **linear regression model** with the following requirements:

  1 - The data to train the model are encoded

  2 - The trained model is encoded

  3 - The model can be used with encoded data which gives an encoded result

Aim of the rest of this talk

- Construct a simple **linear regression model** with the following requirements:

1 - The data to train the model are encoded

2 - The trained model is encoded

3 - The model can be used with encoded data which gives an encoded result