

*Welcome to*

**COMP 9721: Introduction to  
Machine Learning**

Credit hours: 42

Instructor: Dr. Moe Fadaee  
Email: mo.fadaee@georgebrown.ca

School of Con. Ed.  
George Brown College

1

*Introduction to Machine Learning COMP 9721, George Brown College, Moe Fadaee*

## COMP 9721: Introduction to Machine Learning

To know each other:

- Myself
- Yourself



## COMP 9721: Introduction to Machine Learning

Program: **Practical AI and Machine Learning Certificate**

Website: <https://coned.georgebrown.ca/courses-and-certificates/practical-ai-and-machine-learning-certificate>

COMPULSORY COURSES (FOUR):

1. COMP 9721: Introduction to Machine Learning
2. COMP 9723: Machine Learning: Practical Data Analysis
3. COMP 9724: Neural Networks in Artificial Intelligence
4. COMP 9725: Computer Vision and Image Processing

## COMP 9721: Introduction to Machine Learning

### **Course 1: Introduction to Machine Learning (COMP 9721)**

- 1. What is Machine Learning?**
- 2. Programming languages and different environments,**
- 3. Fundamental of statistics**
- 4. Data preparations, preprocessing...**
- 5. Regression and classification**
- 6. Python programming, different libraries and packages**

## COMP 9721: Introduction to Machine Learning

### Course 2: **Machine Learning: Practical Data Analysis (COMP 9723)**

1. What is Machine Learning? Continued
2. Building and evaluating: Regression, Classification and Clustering
3. Polynomial Regression, Support Vector Regression, Decision Tree and Random Forest regression models,
4. Logistic Regression, K-nearest Neighbors, Support Vector Machines (SVM) and Kernel SVM, Random Forest classification, Naïve Bayes classification models,
5. Built K-means clustering and Hierarchical clustering,
6. Python programing, different libraries and packages

## COMP 9721: Introduction to Machine Learning

### **Course 3: Neural Networks in Artificial Intelligence (COMP 9724)**

- 1. What is Machine Learning? Continued**
- 2. Neural Networks**
  - I. Artificial Neural Networks (ANN)
  - II. Convolutional Neural Networks (CNN)
  - III. Recurrent Neural Networks (RNN)
  - IV. ...
- 3. Built, train and evaluate ANN, CNN, RNN on real life problems with real data,**
- 4. Python programing, different libraries and packages**

## COMP 9721: Introduction to Machine Learning

### Course 4: Computer Vision and Image Processing (COMP 9725)

1. What is Machine Learning? Continued
2. Image processing techniques
3. Feature and object detections
4. Face detection/recognition
5. Text detection and extraction
6. Image classification
7. Python programming, different libraries and packages

## COMP 9721: Introduction to Machine Learning

### Course outline:

1. Different approaches of Machine Learning and types of learning algorithms,
2. Jupyter Notebook and Python environments, installation guide,
3. Fundamental concepts of statistics like, probabilities, mean and median, variation and standard deviation, probability density function, data distribution, covariance and correlation, normalization...
4. Data types, data-frames and arrays, random number generators, indexing and slicing, statements and conditionals, loops, function, classes and more in Python,
5. Python important libraries like Numpy, Pandas, Matplotlib, Scikit-Learn, Scipy and more,
6. Data management and pre-processing,
7. Simple linear regression and multiple regression, polynomial regression,
8. Version Control

## COMP 9721: Introduction to Machine Learning

### Course Evaluation:

Assessment Tool:	Description:	Outcome(s) assessed:	Date / Week:	% of Final Grade:
Participation and Team work	Attendance, Role Play & Peer Review	1, 2, 8, 11		10%
Assignment 1		1 to 7, 10, 11	Lesson 5	30%
Assignment 2		1 to 7, 10, 11	Lesson 10	30%
Assignment 3		1 to 7, 10, 11	Lesson: 14	30%
			<b>TOTAL:</b>	<b>100%</b>

Learning through discussion, engagements and active participation

Assignment's presentation

Discuss coding strategies, plan of attacks, and solutions...

## COMP 9721: Introduction to Machine Learning

### Who is this program for:

- If you are interested to find a job in data analysis field
- If you have a job and want to know how AI can help you advance, and do things more efficiently
- If you are just curious about AI and ML and want to know more

## COMP 9721: Introduction to Machine Learning

### Program requirements:

- No prior knowledge in AI is required
- No prior skill in programming with python is required, but is advantageous
- Students need to have access python programming environment
- We are going to use Anaconda

## COMP 9721: Introduction to Machine Learning

### What is AI and Machine Learning:

#### 1. Artificial Intelligence:

*"In computer science, [artificial intelligence \(AI\)](#), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving""*

Wikipedia, Jan 2020

#### 1. Machine Learning (a subdivision of AI):

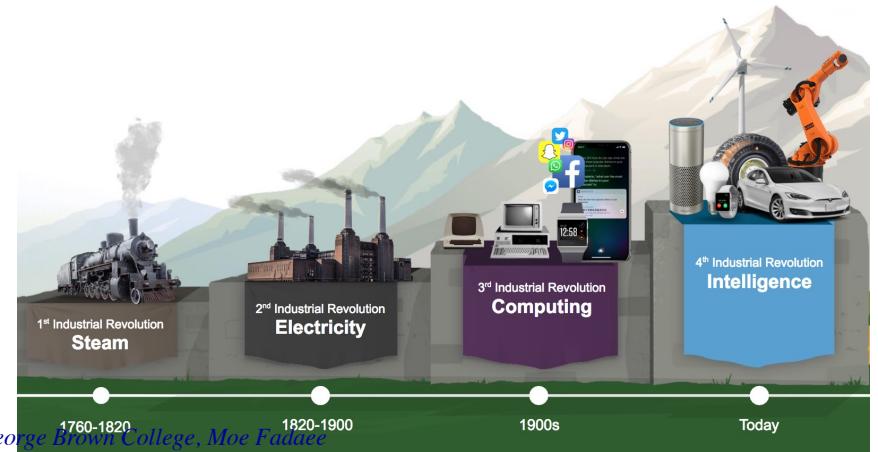
*"[Machine learning \(ML\)](#) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.[1][2] Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task."*

Wikipedia, Jan 2020

**AI is not new!**

**We now have the technology to adopt it!**

1. We have seen AI applications all around us
2. Forth industrial revolution
3. AI is changing the way we live and work
4. AI, just a new technology or a new culture?



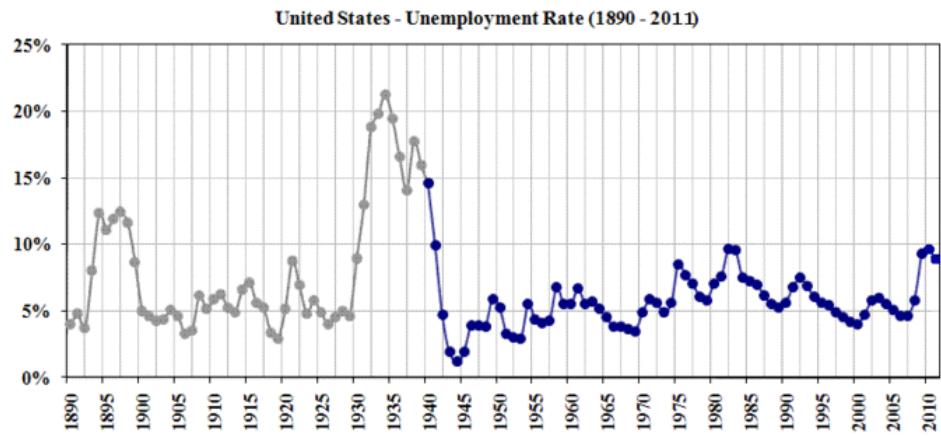
## COMP 9721: Introduction to Machine Learning

Will AI destroy more jobs than it can create or not?

Should we worry?

- 2013 Oxford university study (Carl Benedikt Frey and Michael A. Osborne): 47% of US employment is at risk from automation, and generally technology,
- Recent surveys, claimed only 8% of jobs are at risk,
- AI will create new occupations we can't imagine today. However those jobs are most likely to need new skills,

**Adopt ourselves!**



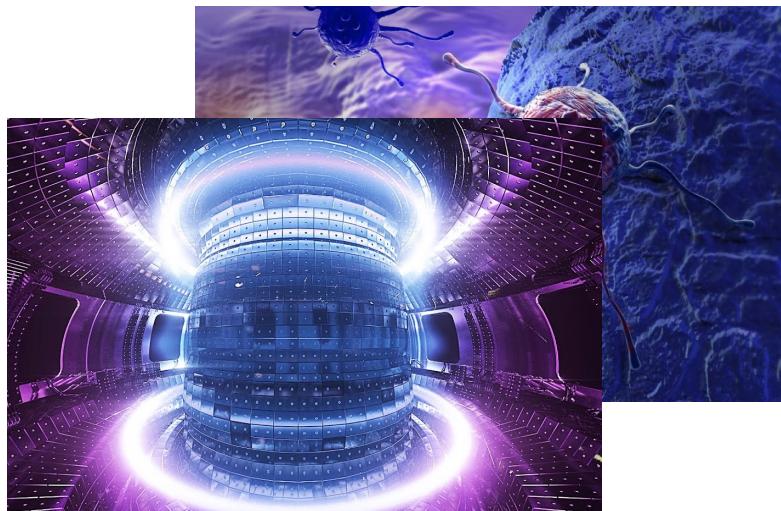
## COMP 9721: Introduction to Machine Learning

AI can play hide and seek! Impressive!

But...



- AI can help us to solve the unsolvable,
  - Cure diseases
  - Harness Fusion
  - Built advance new reactor concepts
  - ...



## COMP 9721: Introduction to Machine Learning

### What is really Machine Learning:

The process that enable **Machine** to **Learn** from **Data**.

1. We need data to train a model,
2. We need a little programming,
3. Machine will do most of the job,
4. Analyse the results after, if needed re-train.

Your model can get as good as the **data** that it trained on!

## COMP 9721: Introduction to Machine Learning

### Machine Learning applications:

Here are some of ML applications:

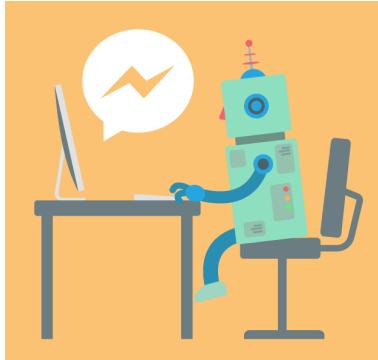
1. Chatbots and Virtual Personal Assistants (agents)
2. Autonomous cars
3. Anomaly detections, e.x detecting fraud transaction in bank accounts
4. Recommenders systems, Netflix, IMDB...
5. Product reviews analysis, AMAZON...
6. Noise filtration
7. Audio to text and text to audio conversion
8. Face and object detection, video surveillance
9. Email spams
10. Search Engines
11. ...



## COMP 9721: Introduction to Machine Learning

### Machine Learning revolutionized industries (or will) :

- I. Here are some of ML applications:
- II. Sale and retail, marketing, business development
- III. Chatbots, revolutionized online services
- IV. Manufacturing, by automation, replacing robots
- V. Healthcare, X-rays and image processing, diagnose diseases (earlier than we do now),
- VI. Educations, (think of AI teachers)
- VII. Real state and construction industry, (smart homes, automate construction)
- VIII. ...



18

## COMP 9721: Introduction to Machine Learning

**Before getting into different type of ML algorithm, lets learn a little about features:**

**Feature (Or variable):**

Is an individual measurable property or characteristic of a phenomenon being measured and observed.

Simply features are those properties of a problem based on which you would like to predict results. A variable is any characteristics, number, or quantity that can be measured or counted.

Like: Number of bedrooms in a house, age, amount of rain, grade in a course, date, color, province, score...

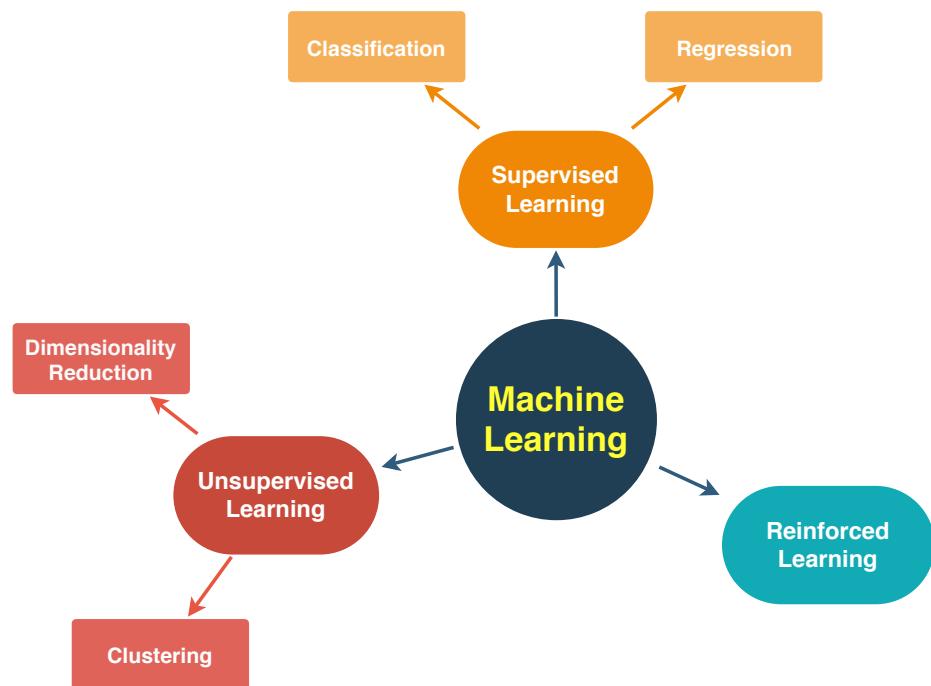
**Tell me more?**

**We will get back to this.**

## COMP 9721: Introduction to Machine Learning

### Machine Learning subcategories:

- I. Supervised Learning
- II. Un-supervised Learning
- III. Reinforced learning
- IV. Others...

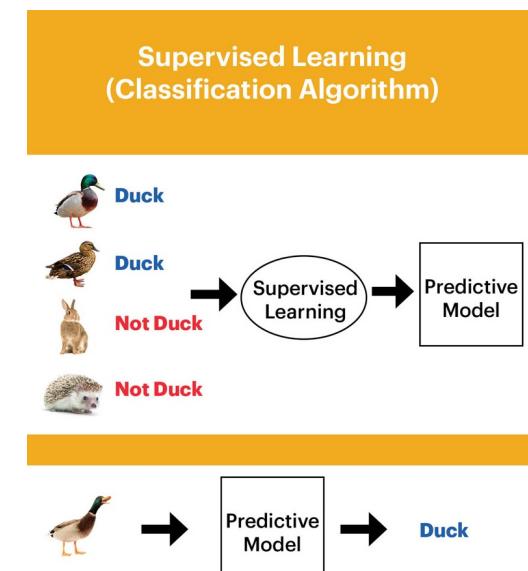
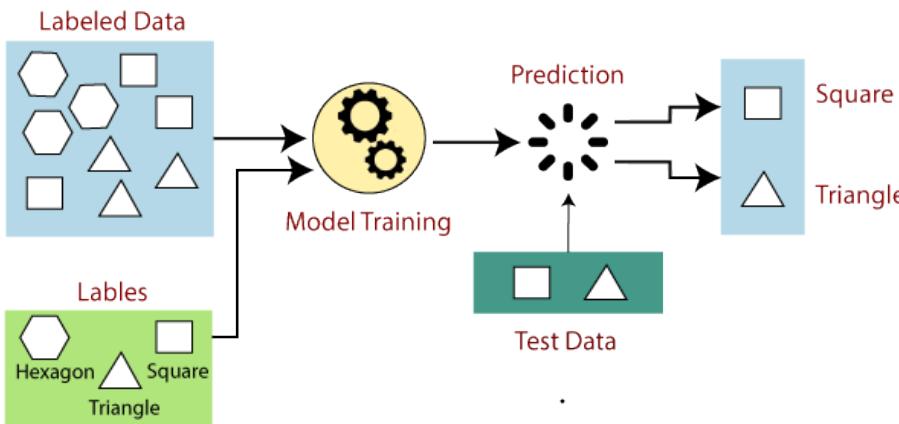


## COMP 9721: Introduction to Machine Learning

### Supervised Learning:

In supervised learning, a computer program is trained with a dataset that is labelled with corresponding output values, and a function will be determined based on this dataset

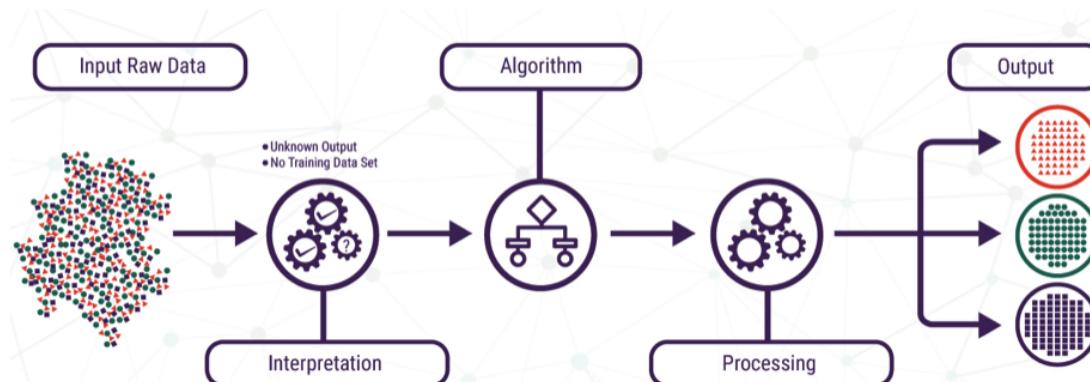
The model will generate an algorithm based on these dataset. After algorithm is developed new un-labelled data can feed into the algorithm/function and answer will be predicted. Supervised Learning can be directly applied to Regression or Classification problems.



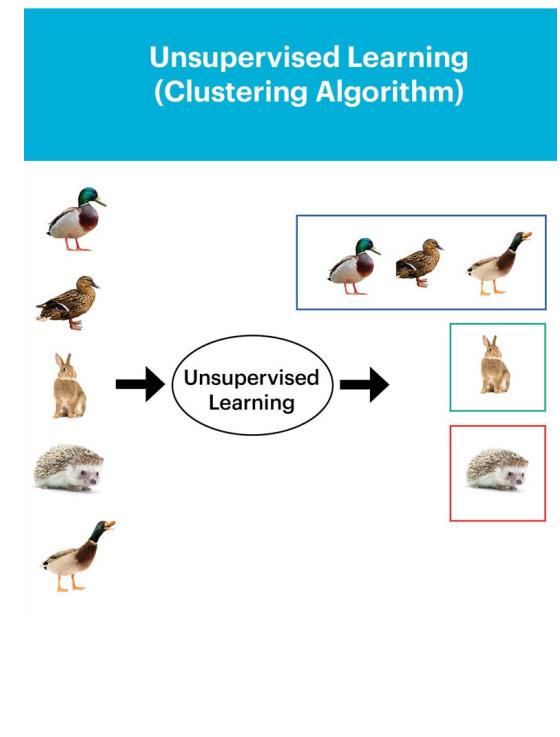
## COMP 9721: Introduction to Machine Learning

### Un-supervised Learning:

Unsupervised machine learning algorithms extract patterns from a dataset without reference to labeled or known outcomes. Unlike supervised machine learning, unsupervised machine learning methods cannot be directly applied to a regression or a classification problem because you have no idea what the values for the output data might be, making it impossible for you to train the algorithm the way you normally would. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).



*Introduction to Machine Learning COMP 9721, George Brown College, Moe Fadaee*



## COMP 9721: Introduction to Machine Learning

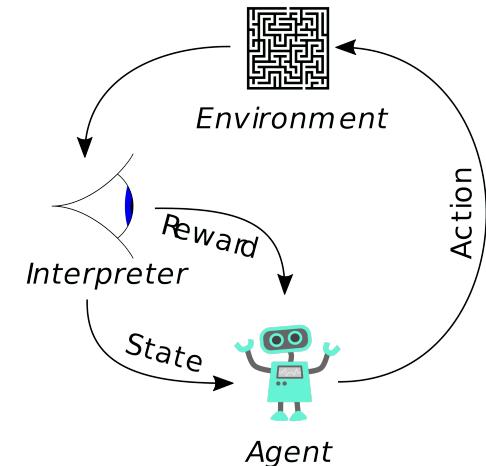
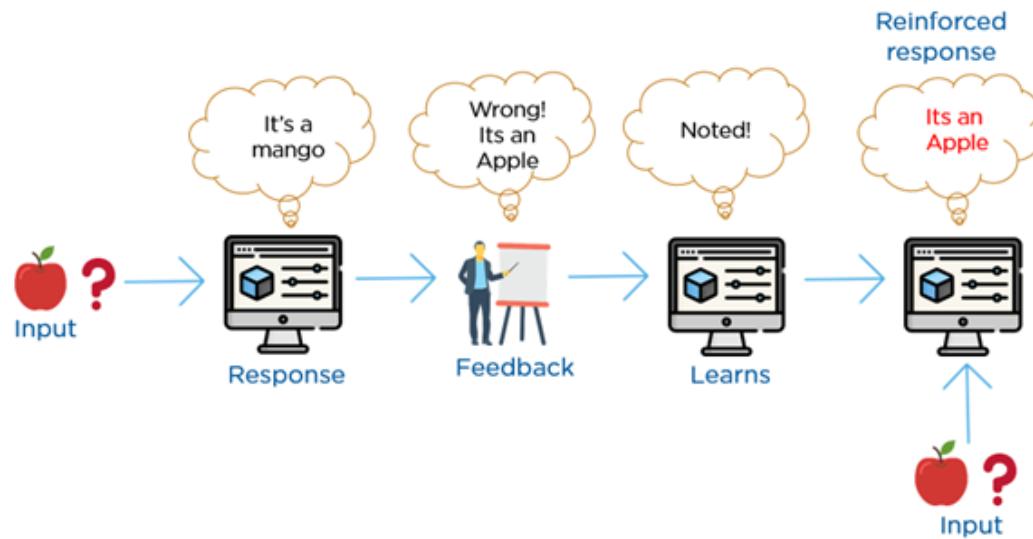
### Reinforced learning:

Reinforced Learning also known as Online learning:

In Reinforced learning a machine learns using trial and error method. Here basically, we give the machine 2 instructions:

1. Try all possible choices.
2. Based on past experiences avoid errors and increase success rate.

Like a human baby learn things by trying different things.



## COMP 9721: Introduction to Machine Learning

### Reinforced learning

RL is based on

- Agent and Environment
- Interaction between the agent and the environment through rewards.
- How software agent should take action in order to maximize its rewards.
- Supervised learning vs. Reinforced learning !?

Applications:

- AI and Robotics
- Gaming
- Industrial automation
- Optimization (Software as well as hardware), Use RL to tune and optimize a NN
- Forecasting
- Product Design
- Text and speech

...

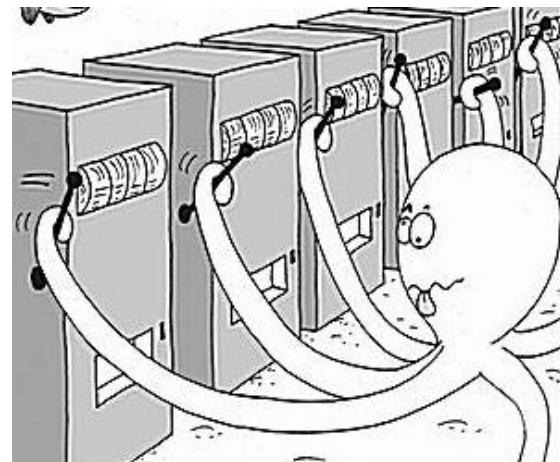
Needs lots of data (lots of trial and error)

Reinforced learning:

- Multi-armed bandit problem,
  - Maximize the reward by deciding about
    - Which machine to play
    - How many times to play
    - Which order to play

...

The Exploration-Exploitation trade off



### Regression and prediction:

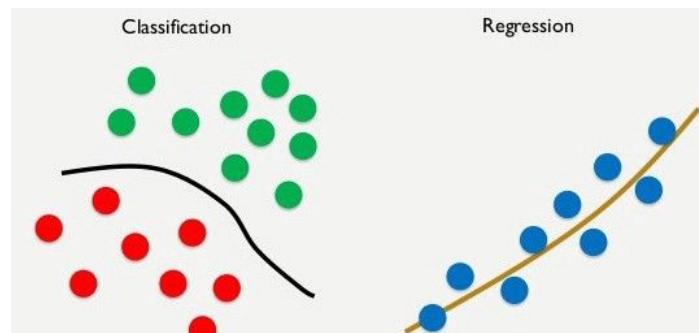
Regression analysis is a set of processes for estimating the relationships between a dependent variable or the mapping function and one (or more) independent variables (also called predictors, covariates or features):

$$y = F(x_1, x_2, \dots)$$

It predicts continuous valued output (works with numeric values).

So, to recap:

- Linear Regression is a machine learning algorithm based on supervised learning.
- It performs a regression task.
- Regression models a target prediction value based on independent variables.

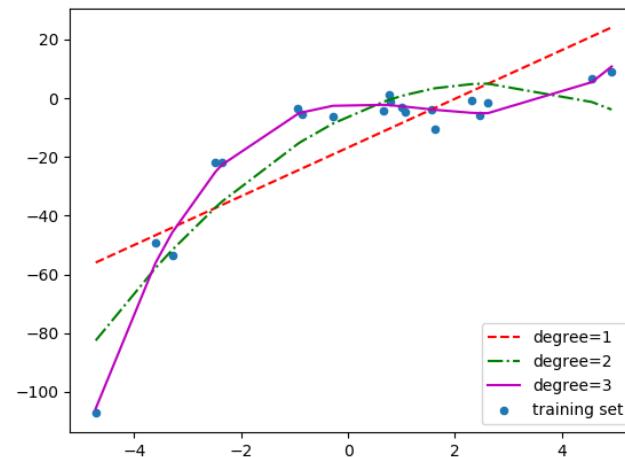
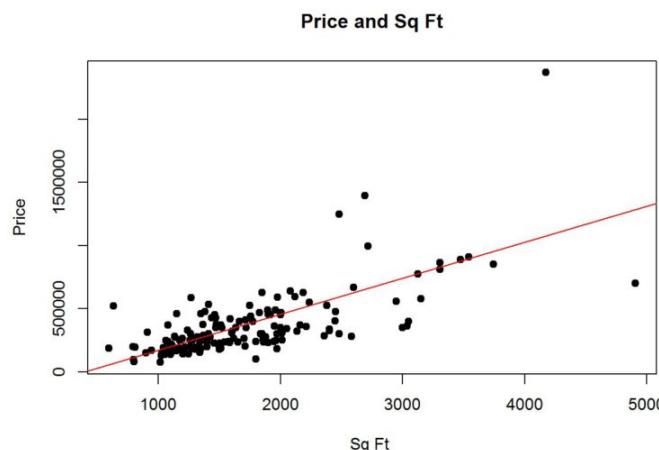


*Introduction to Machine Learning COMP 9721, George Brown College, Moe Fadaee*

## COMP 9721: Introduction to Machine Learning

### Different types of Regression:

- Linear Regression. It is the simplest form of regression. ...
- Polynomial Regression. It is a technique to fit a nonlinear equation by taking polynomial functions of independent variable. ...
- Logistic Regression. In logistic regression, the dependent variable is binary in nature (having two categories).
- Quantile Regression. ...
- Ridge Regression. ...
- Lasso Regression. ...
- ...



27

## COMP 9721: Introduction to Machine Learning

### Regression Applications:

Prediction (e.g. house prices, sale, rate of failure...)

Take care of missing data

Risk assessment

Classification (Logistic regression, e.g. fail or not, vote or not...)

So many others...

## COMP 9721: Introduction to Machine Learning

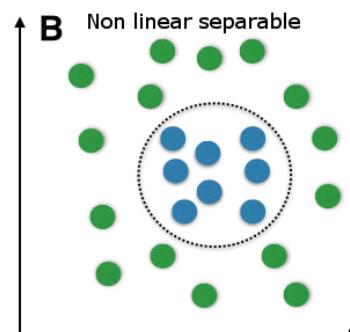
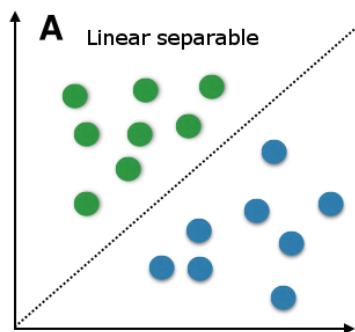
### Classification:

In classification the goal is to identify which set of categories (sub-populations) a new observation belongs to, based on the previous training datasets (data containing observations (or instances) whose category membership is known). It predicts discrete number of values (unlike regression).

As a famous example, an email can be classified to one of two classes: “spam” and “not spam”.

Note that:

- A classification problem requires that examples be classified into one of two or more classes.
- A classification can have real-valued or discrete input variables.
- Classification with two classes is often called a two-class or binary classification problem.
- Classification with more than two classes is often called a multi-class classification problem.
- A problem where an example is assigned multiple classes is called a multi-label classification problem.



29

## COMP 9721: Introduction to Machine Learning

### Classification Applications:

Text Classification (e.g. emails, reviews...)

Image classification

Sentiment Analysis, (e.g. attitude or opinion of the speaker or the writer)

News classification

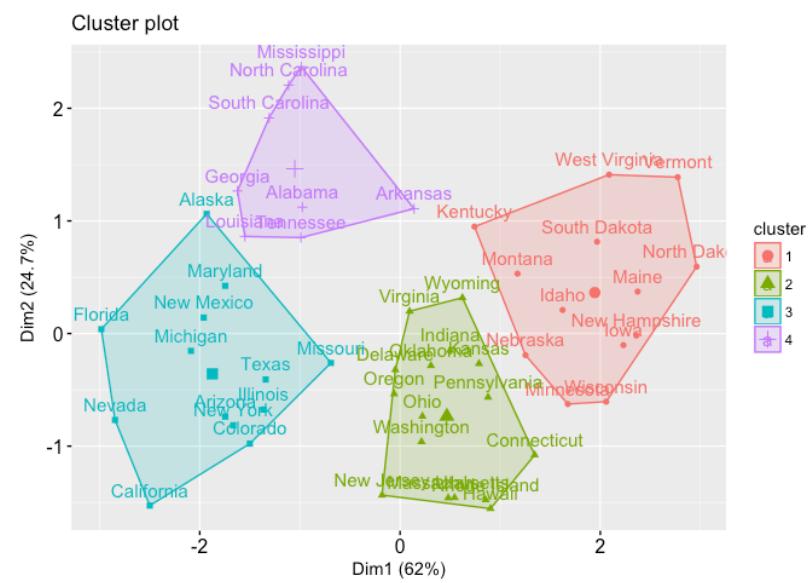
## COMP 9721: Introduction to Machine Learning

### Clustering:

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

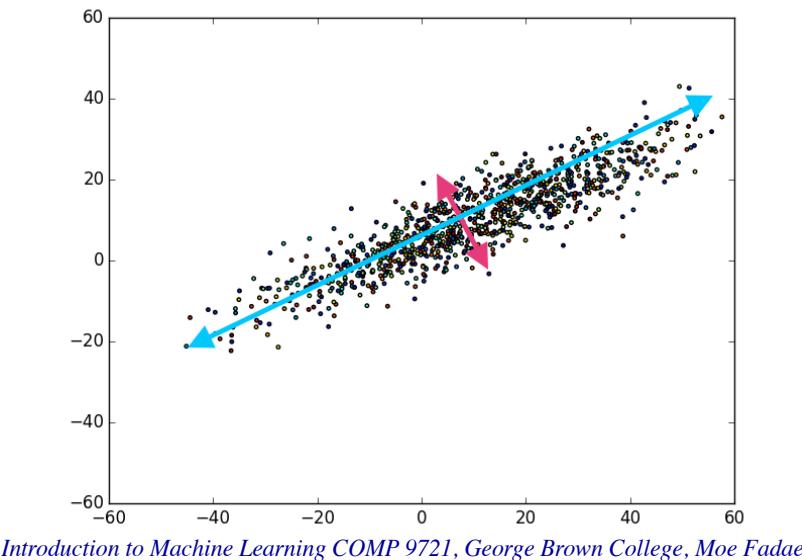
### Applications of Cluster Analysis

- Market research, pattern recognition, data analysis, and image processing.
- Characterize their customer groups based on the purchasing patterns.
- It can be used to derive plant and animal taxonomies, categorize genes with similar functionalities
- Clustering also helps in classifying documents on the web for information discovery.
- Detection of credit card fraud and similar applications
- ...



## Dimensionality Reduction:

Reduces number of variables (features or factors) by obtaining a set of principle variables. When number of variables are high, it is hard to visualize the data. Moreover, most of the time these variables are correlated and therefore consider to be redundant. As an example consider house prices as a function of number of rooms and number of bathroom.



## Dimensionality Reduction:

Dimensionality reduction can be divided into:

- Feature selection: is the process of selecting the most important and relevant features (variables) from the whole set of features. E.g to pick number of rooms and eliminate number of bathrooms, when training a model to predict house prices.
- Feature extraction: To actually reduce the number of features by extracting new features. E.g. to extract a new feature that is a function of No of rooms and bathrooms, i.e. the new feature has both of the old features (maybe No of rooms X No of bathrooms).

## COMP 9721: Introduction to Machine Learning

### Exercises:

What kind of ML algorithm you would use for this problem:

A bank need a machine learning algorithm to scan all transactions on then account and identify if a transaction is legit or fraud.

What are the features?

What is the proper ML algorithm to solve this problem?

## COMP 9721: Introduction to Machine Learning

### Exercises:

What kind of ML algorithm you would use for this problem:

A construction company or a car part manufacturer wants to find out if its employee wearing protective equipment's or not.

What are the features?

What is the proper ML algorithm to solve this problem?



## COMP 9721: Introduction to Machine Learning

### Exercises:

What kind of ML algorithm you would use for this problem:

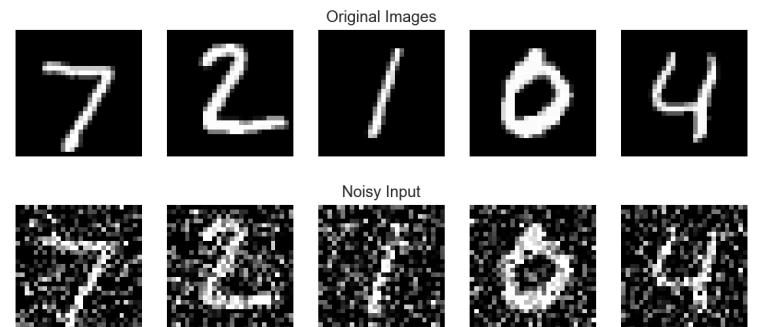
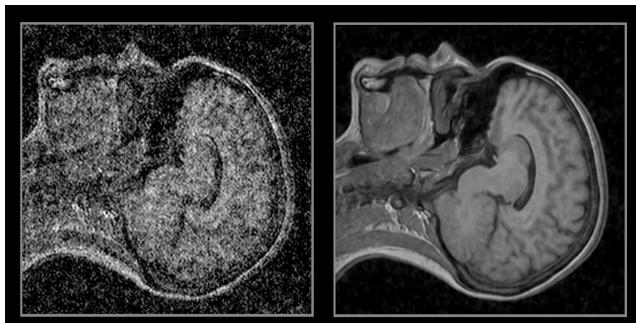
An company active in art need to train a ML model that denoise images,

Or

A medical company that need to denoise scans

What are the features?

What is the proper ML algorithm to solve this problem?



## COMP 9721: Introduction to Machine Learning

### Exercises:

What kind of ML algorithm you would use for this problem:

A financial institute wants to build a ML model that predict stock market price for the next day,

Or

Predict electricity price for next day

What are the features?

What is the proper ML algorithm to solve this problem?



37

## COMP 9721: Introduction to Machine Learning

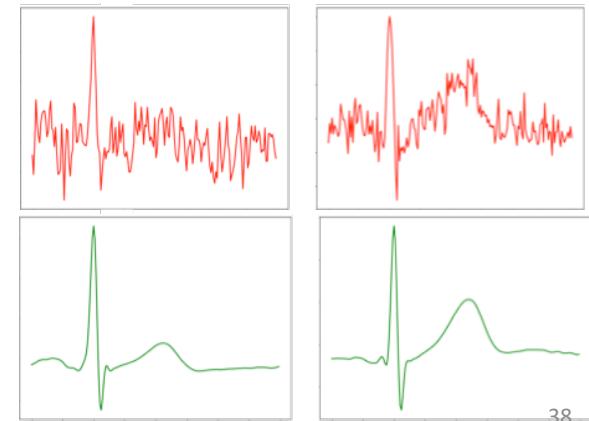
### Exercises:

What kind of ML algorithm you would use for this problem:

A company that built medical device want to built an ML model that collect medical data like (heart rate, blood pressure...) and filter the noise,

What are the features?

What is the proper ML algorithm to solve this problem?



## COMP 9721: Introduction to Machine Learning

### Exercises:

What kind of ML algorithm you would use for this problem:

A company that built car engine, want to train a model to tell if anything is wrong with the engine or not, if yes, most likely what is the part that causes the problem?

What are the features?

What is the proper ML algorithm to solve this problem?

## COMP 9721: Introduction to Machine Learning

Look at some of things we have achieved using AI and ML:

Movie by Benjamin:

<https://www.youtube.com/watch?v=LY7x2lhqjmc>

<https://www.youtube.com/watch?v=vUgUeFu2Dcw>

Hide and sick agents:

<https://www.youtube.com/watch?v=kopoLzvh5jY>

Autonomous cars:

<https://www.youtube.com/watch?v=l3ELVACR2VY>

This is from Renault, but there are so many others out there.

So many other cool applications...

<https://www.youtube.com/watch?v=nOnsbd7rdhc>

And still more...

# Statistics



## COMP 9721: Introduction to Machine Learning

**Before getting into different type of ML algorithm, lets learn a little about features:**

**Feature (Or variable):**

Is an individual measurable property or characteristic of a phenomenon being measured and observed.

Simply features are those properties of a problem based on which you would like to predict results. A variable is any characteristics, number, or quantity that can be measured or counted.

Like: Number of bedrooms in a house, age, amount of rain, grade in a course, date, color, province, score...

## COMP 9721: Introduction to Machine Learning

### Features and variable types (Data type):

- Numeric variables are those that have a measurable quantity as a number. Numeric variables are **quantitative variables**.

Examples?

Numeric variables are either continuous or discrete:

A continuous variable can take any value between a certain set of real numbers (between 0 and 1).

Examples?

A discrete variable can take a value based on a count from a set of distinct whole values.

Examples?

- Categorical variables have values that describe a 'quality' or 'characteristic' of a data unit. Therefore, categorical variables are **qualitative variables** and will be represented by a non-numeric value.

Categorical variables are ordinal or nominal:

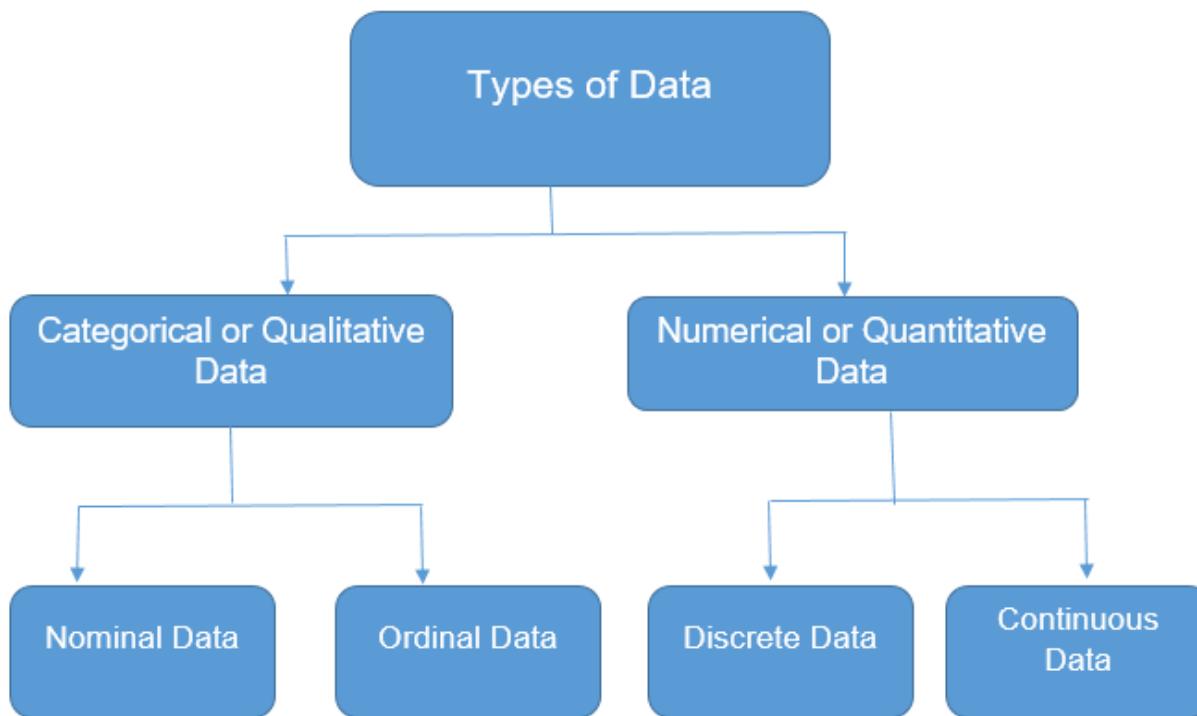
An ordinal variable can take a value that can be logically ordered or ranked.

Examples: quality of a restaurant: poor, good, excellent, others?

A nominal variable can take a value that is not able to be organised in a logical sequence.

Examples hair color, others?

Data:





Programming Languages:

Main programming language in this program is Python (Not the snake!).

Created by Guido van Rossum in 1991.

[www.python.org](http://www.python.org)

It received a huge attention recently.

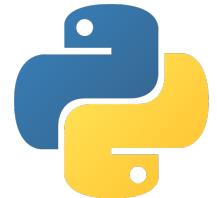
Python is relatively slow:

1. It is a high level programming language, code is more understandable and it is similar to how humans thinks (You do less, computer do more).
2. It is dynamic, size and type of variables wont be declare at the beginning.

Pros:

1. It is productive and easy to understand. It requires less time, effort, and lines of code to perform the same task in other languages.
2. There are several packages and libraries that we can use. We will explore them in this course.





**Installing Python:**

It is highly recommend to install Anaconda (Again not the snake!).

Anaconda is a open source platform for python and R languages. It can be downloaded from

<https://www.anaconda.com/distribution/>

With anaconda already comes with so many packages,