

# Predictive Modeling and Data Mining

## Week 1- Introduction

Mohammad Esmalifalak

January 2019





## Work load and Evaluation

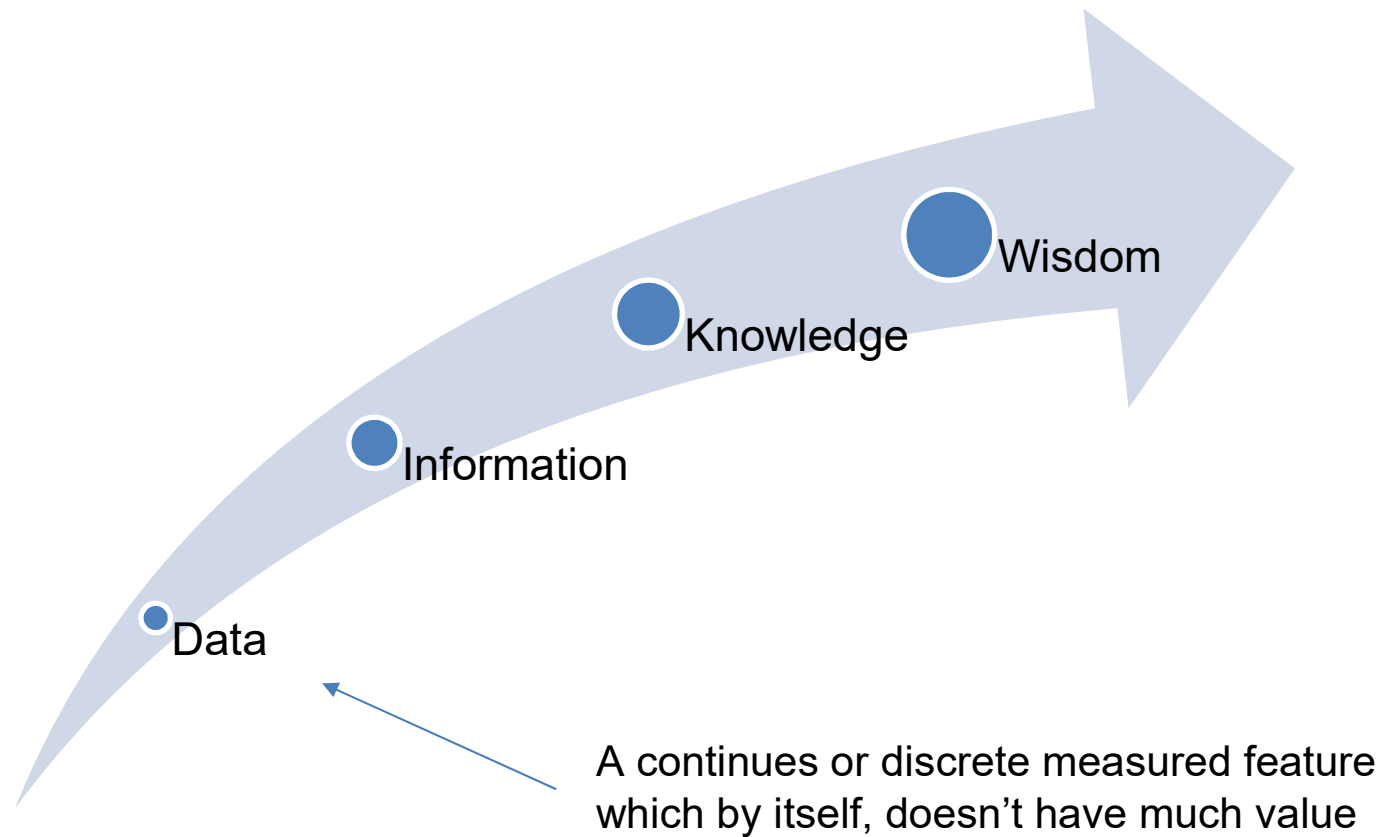
- A mix between theory and lab work
- 10 Marked labs
- Lab Assignments 65% Labs (See deadlines)
- Report + presentation 35% (See deadlines)



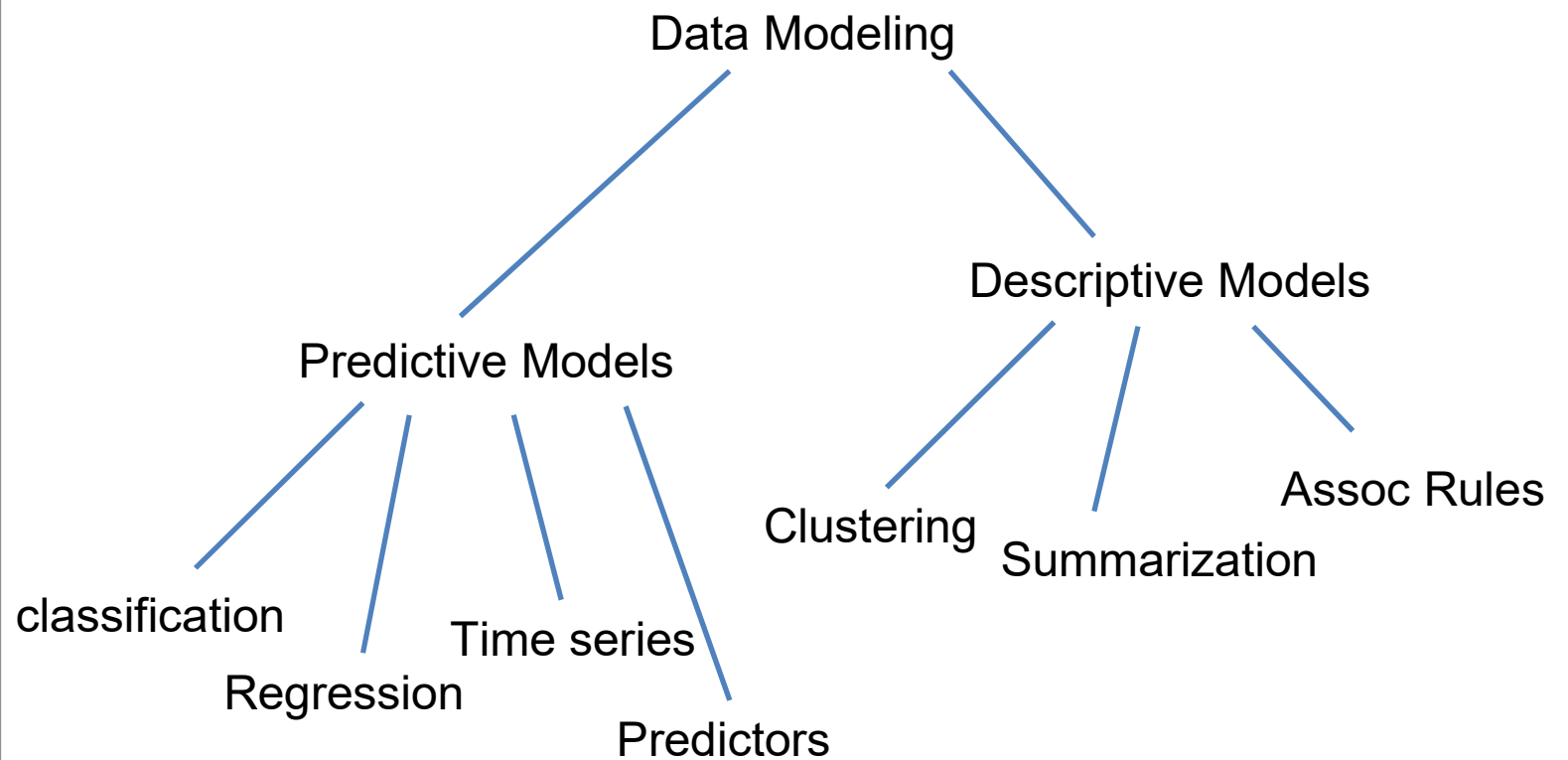
## Course Objective

- The course will introduce predictive modelling techniques as well as related statistical and visualization tools for data mining.
- The course will cover common machine learning techniques that are focused on predictive outcomes.
- Students will learn how to evaluate the performance of the prediction models and how to improve them through time.

# What's Data: Definitions

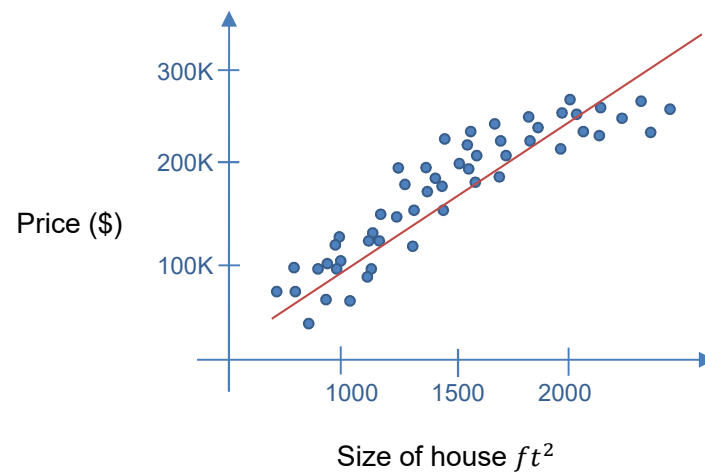


# Data Modeling



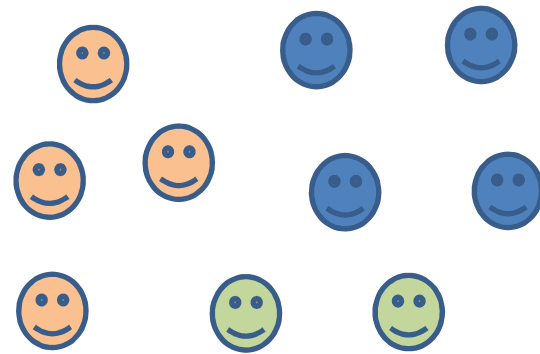
# Data Modeling

## Supervised Learning



We know the TRUE output and algorithm learns from that.

## Un-supervised Learning

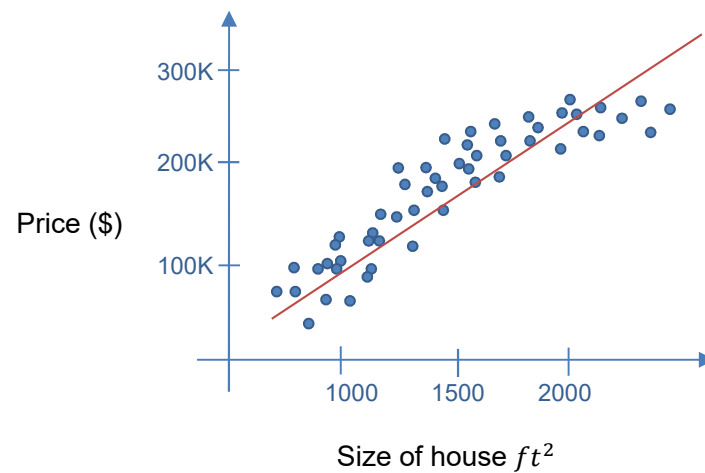


Market Segmentation (Clustering)

We don't know the TRUE output  
(number of clusters, clustering rules, etc. )

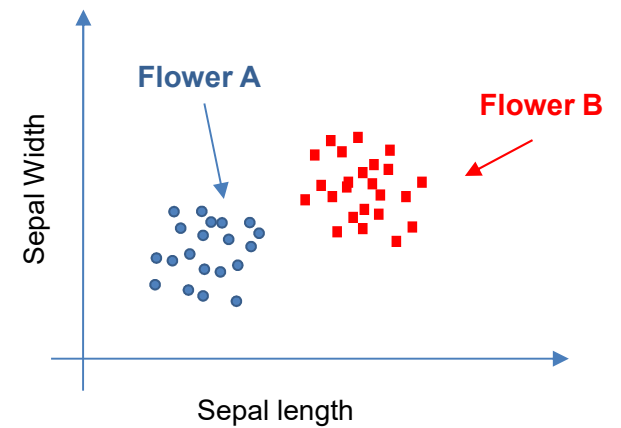
# Data Modeling

## Regression



Output label is continuous

## Clustering



Output label is discrete (2 or more classes)

# Predictive Modeling in Python

Data Cleaning  
with **Pandas**



Array Operation  
with **Numpy**



Predictive  
Algorithms with  
**Sklearn**





Downloaded from <http://ajph.org/> on November 10, 2014



# Course Schedule

Week	#	Lesson/Module Title	Evaluation	Evaluation Due Dates	Delivery Method
Jan 16	1	Course Introduction Introduction to Predictive Modelling	Lab #1: Lab Activity: Data Science Tools	Jan 22; 11:59 PM	In-class lecture / Lab Demonstration
Jan 23	2	Preparing Data and Feature Engineering	Lab #2: Data Cleaning	Jan 29; 11:59 PM	In-class lecture / lab Demonstration
Jan 30	3	Linear Regression & Linear Regression with Variable Selection	Lab #3: Linear Regression with Variable Selection	Feb 05; 11:59 PM	In-class lecture / lab demonstration
Feb 06	4	Classification: Logistic Regression	Lab #4: Logistic Regression	Feb 12; 11:59 PM	In-class lecture / lab demonstration
Feb 13	5	Hyper-parameters, Regularization and Cross-Validation	Lab #5: Improving model performance with regularization	Feb 19; 11:59 PM	In-class lecture / lab demonstration

# Course Schedule

Week	#	Lesson/Module Title	Evaluation	Evaluation Due Dates	Delivery Method
Feb 20	6	Classification : Decision Trees	Lab #6: Decision trees	Feb 26 ; 11:59 PM	In-class lecture / lab demonstration
Feb 27	7	Simulation Modeling	Lab #7: Marketing Simulation	Mar 05; 11:59 PM	In-class lecture / lab demonstration
Mar 06	8	Dimensionality Reduction	Lab #8 Principle Component Analysis	Mar 12; 11:59 PM	In-class lecture / lab demonstration
Mar 13	9	Clustering: K-mean	Lab #9: K-means	Mar 19; 11:59 PM	In-class lecture / lab demonstration
Mar 20	10	Clustering: Shift mean	Lab #10: Shift mean	Mar 26; 11:59 PM	In-class lecture / lab demonstration
Mar 27	11	Spatial-Temporal Analytics	Lab Activity; Examples of Spatial-Temporal Analytics	No submission	In-class
April 03	12	Project Presentations Class Wrap up			In-class



## In this course:

- We will go through various predictive methods and algorithms.
- Programming language will be python.



© All rights reserved. All content within our courses, such as this video, is protected by copyright and is owned by the course author or unless otherwise stated. Third party copyrighted materials (for example, images and text) have either been licensed for use in any given course, or have been copied under an exception or limitation in Canadian Copyright law. For further information, please contact the McMaster University Centre for Continuing Education [ccecrsdv@mcmaster.ca](mailto:ccecrsdv@mcmaster.ca).