Predictive Modeling and Data Mining

Week 1- Introduction

Mohammad Esmalifalak

September 2020

McMaster University
CENTRE FOR CONTINUING EDUCATION

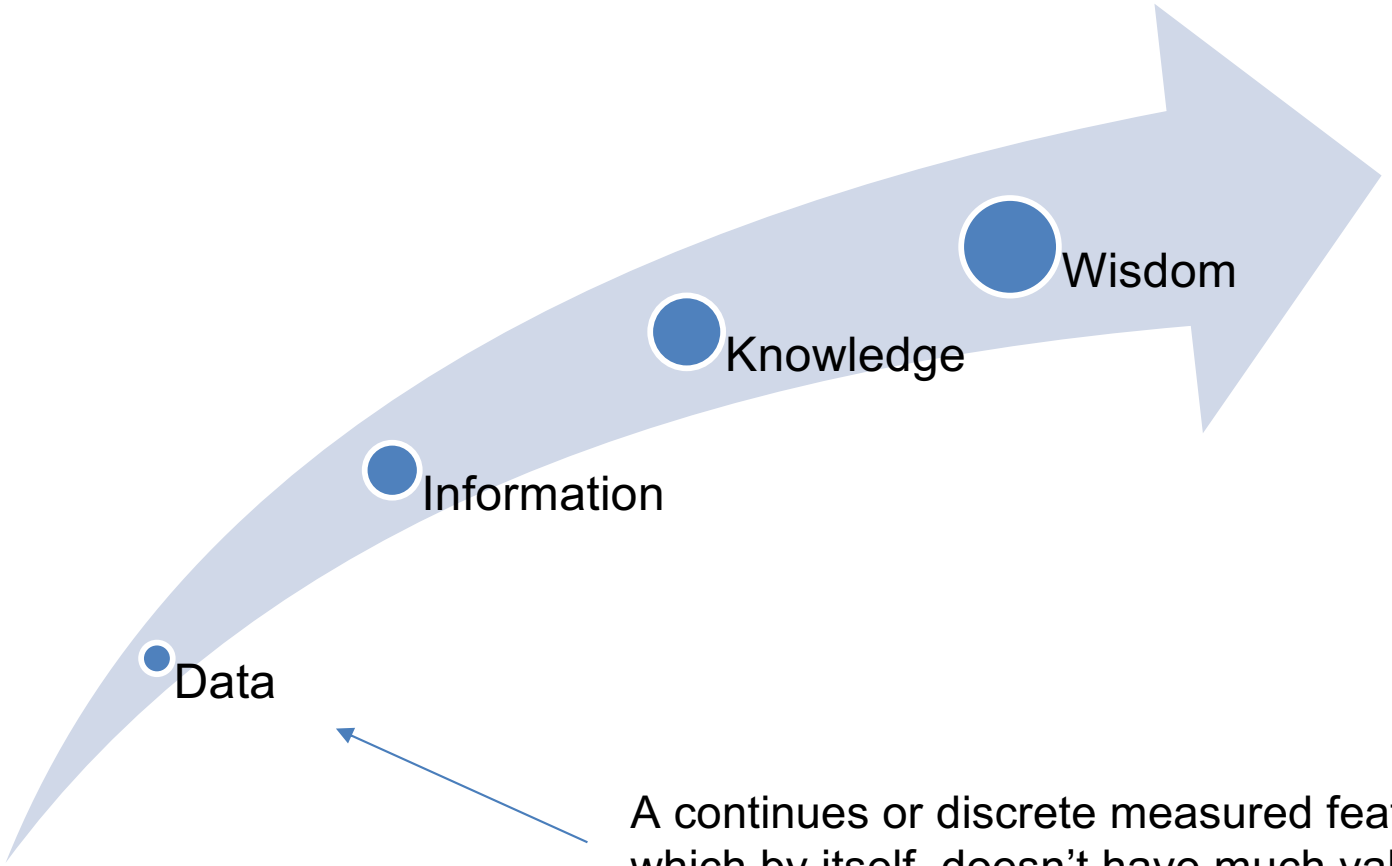
Work load and Evaluation

- A mix between theory and lab work
- 10 Marked labs
- Lab Assignments 50% Labs (See deadlines)
- Report + presentation 50% (See deadlines)

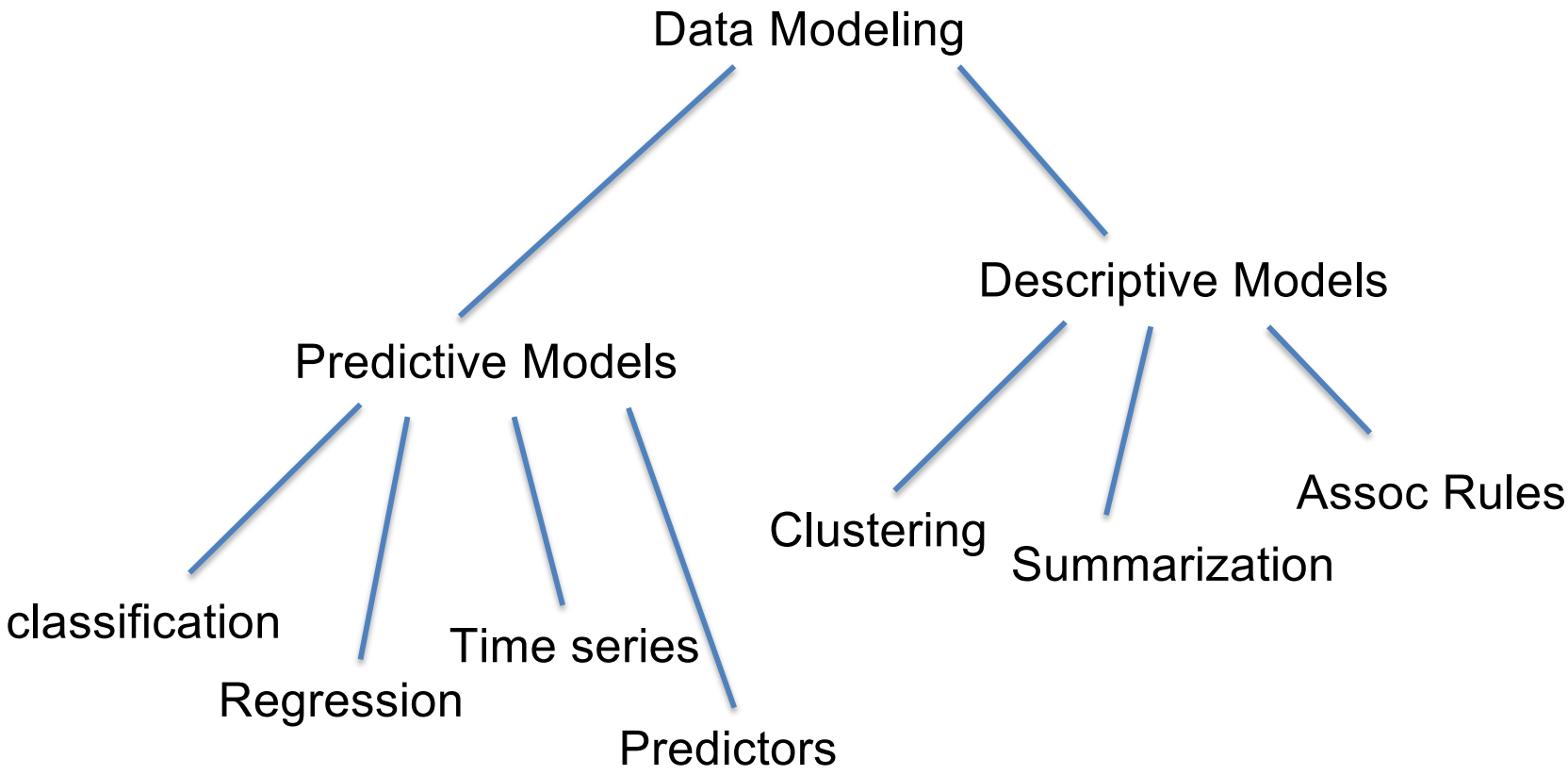
Course Objective

- 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

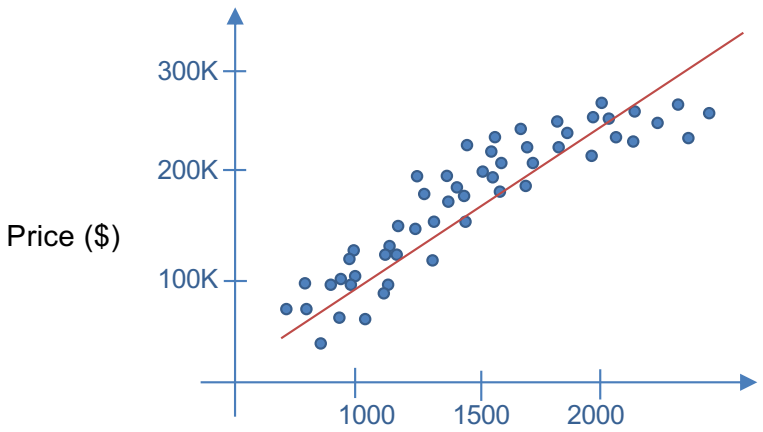


Data Modeling



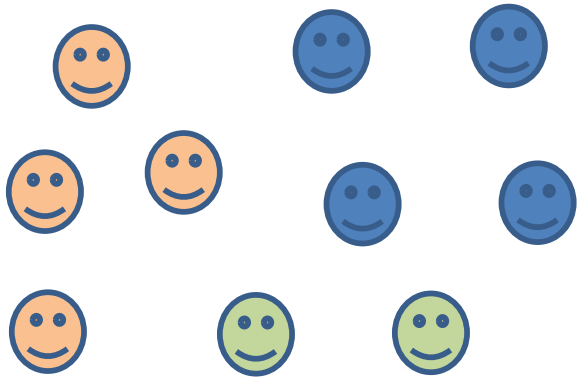
Data Modeling

Supervised Learning



We know the TRUE output and algorithm learns from that.

Un-supervised Learning

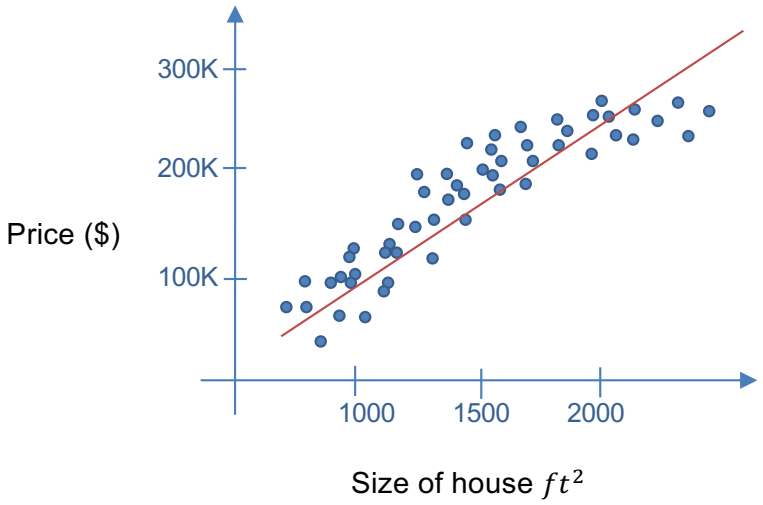


Market Segmentation (Clustering)

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

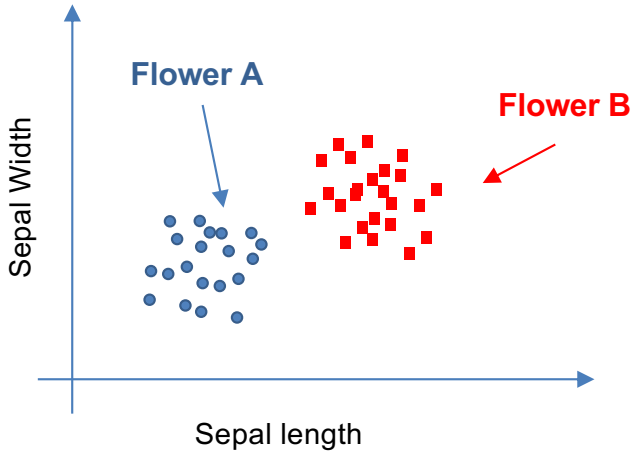
Data Modeling

Regression



Output label is continues

Clustering



Output label is discrete (2 or more classes)

Predictive Modeling in Python

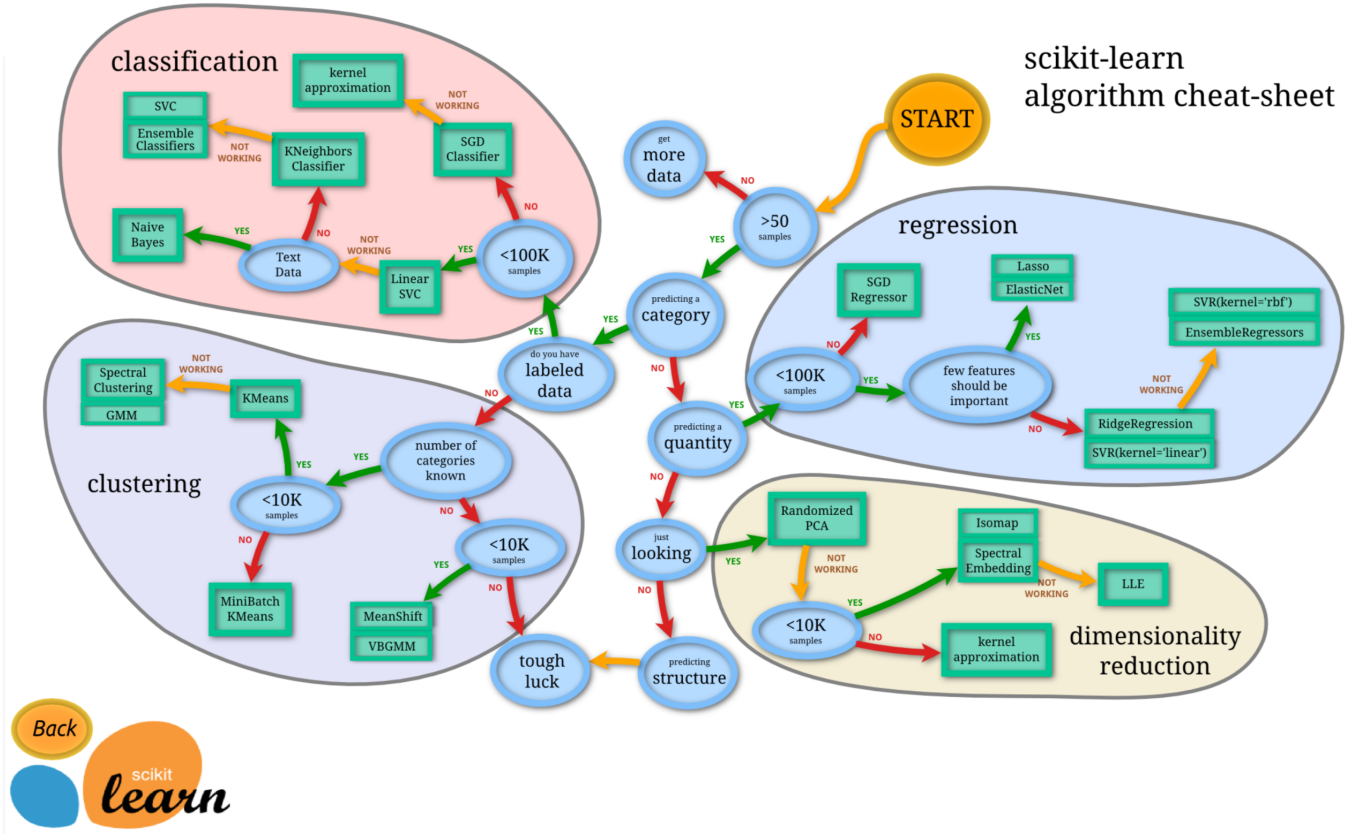
Data Cleaning with **Pandas** → Array Operation with **Numpy** → Predictive Aglorithms with **Sklearn**

NumPy

matplotlib

scikit learn

Scikit-Learn Cheat sheet



Ref: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

Course Schedule

Start Date: September 20, 2020 | **End Date:** December 13, 2020 | All times Eastern

Dates	#	Module	Evaluation	Weight	Due Date	Delivery Method
September 20	1	Course Introduction Introduction to Predictive Modelling	Lab #1: Lab Activity: Data Science Tools	%	-	Electronic Submission
September 27	2	Preparing Data and Feature Engineering	Lab #2: Data Cleaning	5%	October 04	Electronic Submission
October 04	3	Linear Regression & Linear Regression with Variable Selection	Lab #3: Linear Regression with Variable Selection	5%	October 11	Electronic Submission
October 11	4	Classification: Logistic Regression	Lab #4: Logistic Regression	5%	October 18	Electronic Submission
October 18	5	Hyper-parameters, Regularization and Cross-Validation	Lab #5: Improving model performance with regularization	5%	October 25	Electronic Submission
October 25	6	Classification: Decision Trees	Lab #6: Decision trees	5%	November 01	Electronic Submission
November 01	7	Simulation Modeling	Lab #7: Marketing Simulation	5%	November 08	Electronic Submission
November 08	8	Dimensionality Reduction	Lab #8 Principle Component Analysis	5%	November 15	Electronic Submission
November 15	9	Clustering: K-mean	Lab #9: K-means	5%	November 22	Electronic Submission
November 22	10	Clustering: Shift mean	Lab #10: Shift mean	5%	November 29	Electronic Submission
November 29	11	Spatial-Temporal Analytics	Lab Activity; Examples of Spatial-Temporal Analytics	5%	December 06	Electronic Submission
December 06	12	Putting Predictive models into production	-		December 13	
December 13	13	Project Presentations Class Wrap up		50%		

In this course:

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

  https://www.w3schools.com/python/python_reference.asp

McMaster University

CENTRE FOR CONTINUING EDUCATION