Lecture # 1 INTRODUCTION - Data driven modeling

1.0 Introduction

Parameteric vs data driven model.

2.0 Machine Learning workflow

3.0 Source of data

3.1 Synthetic data

3.2 Field data

4.0 Summary

Lecture#1 Data- Driven Machine Learning Modeling

Brief introduction

This part of the Course is data-driven Modeling. (Machine Learning (ML) - Modeling) The lectures present the concepts how ML modeling works? and practical laboratory exercises.

Briefly,

Lecture 1: presents introduction about ML modeling and workflows that we are going to work with.

Lecture 2: Presents the concepts of the first part of ML work flow. That is Feature Engineering /or Data preprocessing

→ Then, Laboratory Practices

Lecture 3: The Mathematics / statistics of How ML modeling and Model performace analysis Conducted / works?

- Here Linear / Poly / Mutivariable Regression will be analyzed.

→ practical Laboratory works

Lecture 4 The Mathematics of how ANN Modeling works?

→ Then, Practical Laboratory works.

Machine Learning Modelling

1.0 Introduction

In this section, we will review the parameteric model and machine learning modelling.

For the machine learning modeling, we will focus on REGRESSION

As you know, for any design analysis works, we need to measure the material properties such as mechanical, physical, electrical, magnetic and other...

Measurment, in general takes time and costy. Moreover, situations may not allow to perform measurment.

Therefore, scientists and Engineers are trying to develope mathematical models to describe the given process / estimate materal properties based on a measured other parameters. However, the ACCURACY of the model should be verified by the measurment dataset

1.1 Mathematical / Parameteric model?

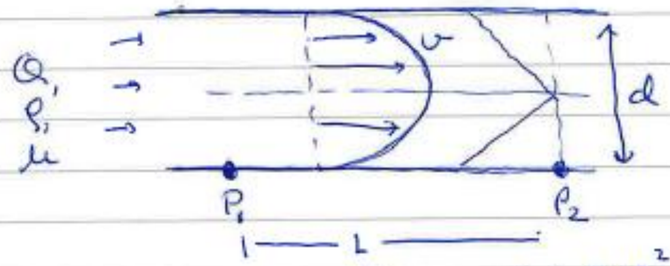
mathematical models are developed based on GOVERNIG LAWS: Example

- Conservation of mass
- Conservation of momentum, Energy..

Examples of Parameteric / Physics models.

Example #1 Hydraulics model.

- Hydraulics model is used to predict pressure loss between two points as fluid flows.



$$\Delta P / L = \frac{2 f \cdot \rho \cdot v^2}{d}$$

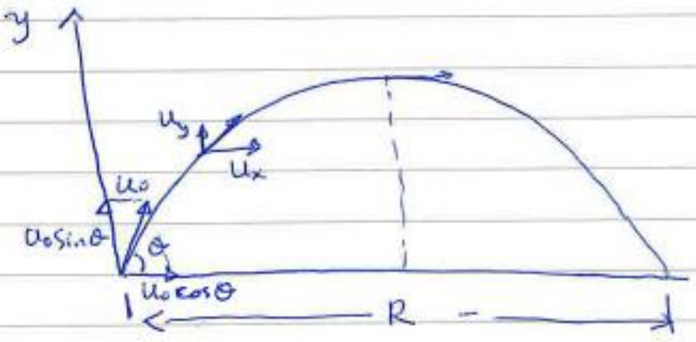
$$f = \frac{16}{Re}$$

$$Re = \frac{\rho \cdot v d}{\mu}, \quad v = Q/A$$

Example 2 Projectile motion

Assume no wind / no drag

$$y = U_0 \sin\theta \cdot t - \frac{1}{2} g t^2$$



- Total time of flight $T = \frac{2 U_0 \sin\theta}{g}$
- Range $R = \frac{2 U_0 \sin\theta \cdot \cos\theta}{g}$

Examples

(3) Simple harmonic motion / vibration

$x = a \cdot \sin(\omega_n t + \varphi)$

(4) Radio active decay

$N(t) = N_0 \cdot e^{-\alpha t}$

(5) Population growth / extinction

$P(t) = P_0 e^{\beta t}$

$\beta(+) \rightarrow$ growth

$\beta(-)$ – extinction

(6) Rectilinear motion

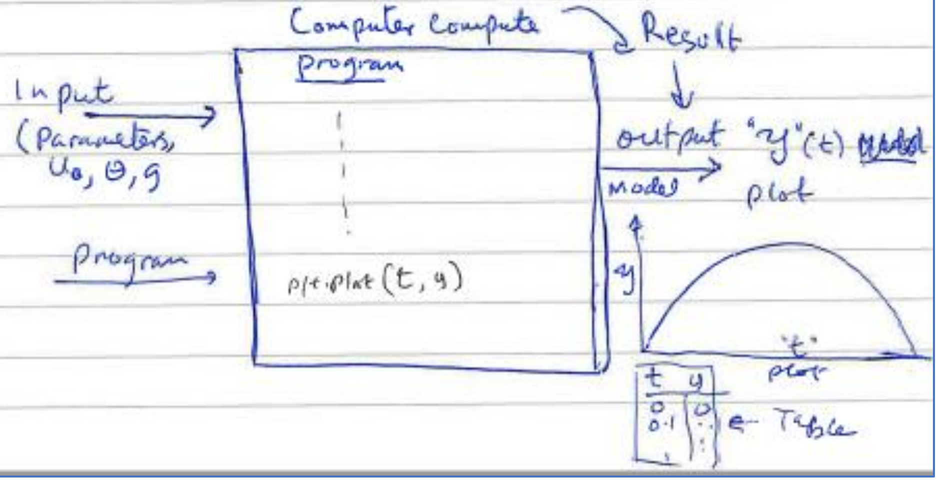
→ constant acceleration

$U_f = U_0 + at$

$S = U_0 t + \frac{1}{2} a t^2$

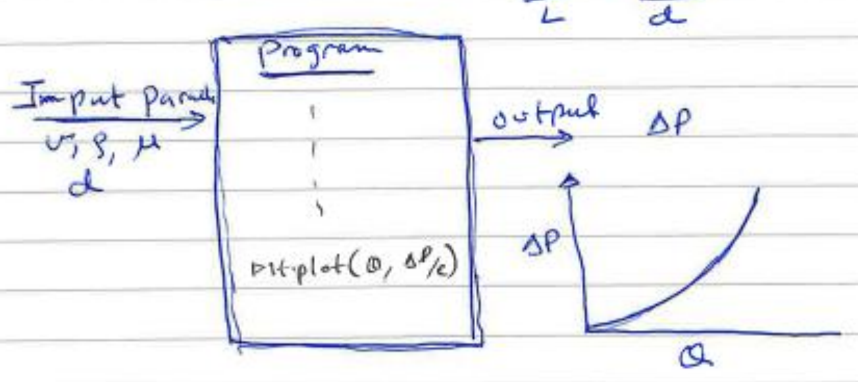
→ To compute parametric model output, we use the traditional programing, by writing the model explicitly and resulting the output, with plot / numeric tabular

Example 1 Projectile motion programing to give $y(t)$



Example 2. To Compute hydraulics models pressure loss

$$\frac{\Delta P}{L} = \frac{2 f \rho v^2}{d}$$



To check the accuracy of the model prediction, we need to compare the model result with a measured data

If there is a discrepancy, we need to Calibrate the model since the physics Model does not capture all the physical phenomena.

The Calibrated model will be

$$\frac{\Delta P}{L} = c \cdot \frac{2 f \rho v^2}{d}$$

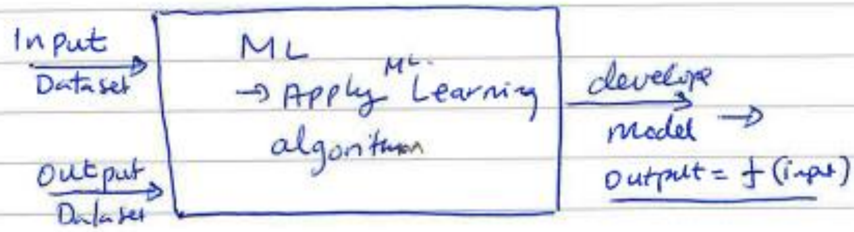
In the calibration factor, c, there are hidden properties

1.2 Machine Learning model?

ML modeling does not apply physical laws, but it generates model by Learning input and output data set

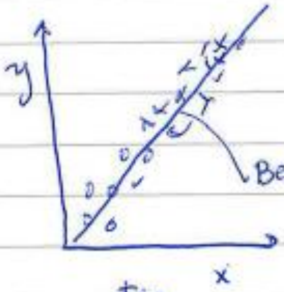
- It is called Data-Driven Modelling

- ML model develope model that 'best fit' the output with input.



ML Learning algorithms are based on the trends between input-output

Example: If the data spread looks like this (Fig below), the trend is Linear and hence, the mapping function that best-suits for the data is Linear

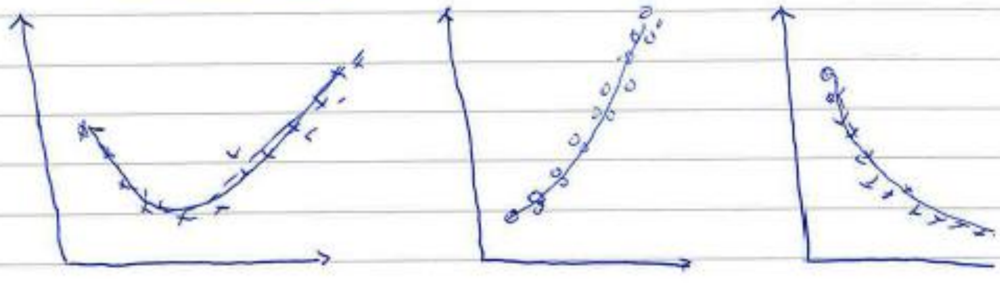


Best-fit $y = \beta_0 + \beta_1 X$

Fig

The ML algorithm will process and find the optimized β_0 and β_1 that gives the "best-fit" model

More examples

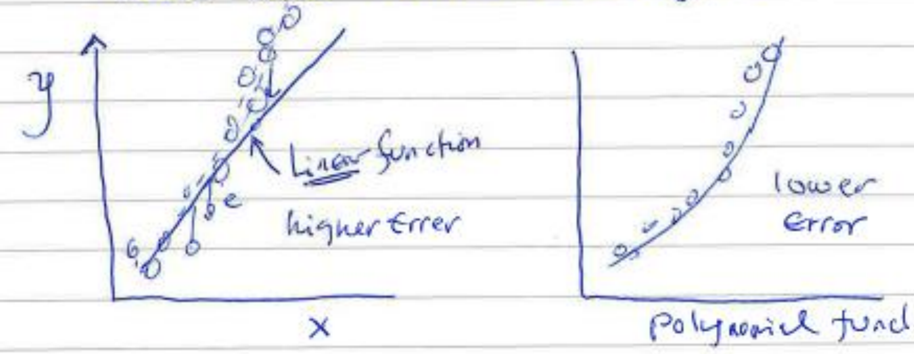


Polynomial mapping function | * Exponential mapping funct * Power law | Exponential mapping funct

- To develope M-L model, DATA is a must
- The QUALITY of data most be CLEAN
- Based on the trend of the data, the right / approximate mapping function Should be SELECTED

Example

If the data trend behaves as a POLYNOMIAL type, donot select a linear mapping function since the residual Error between the data and the model is higher



In this class, we will learn Regression modelling

(I) Linear Regression

(a) Simple Linear regressn $\rightarrow$ $y = \beta_0 + \beta_1 x$

(2) Polynomial regren $\rightarrow$ $y = \beta_0 + \beta_1 x + \cdots \beta_n x_n^{n}$

(3) Multivariable reg $\rightarrow$ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n$

$x_1, x_2, x_3 \cdots x_n$ = input Feature

(~~II~~) The above three mapping functions are called Linear Regression Even though the polynomial contains non-linear term, the coefficients are linear. Therefore, it is under Linear reg. Catagory

(II) Non-Linear Regression.

The function contain non-linear exponent

(a) Exponential $\rightarrow$ $y = ae^{\pm \beta x} + c$

(b) Power $y = ax^{bx} + c$
$(b \in \mathbb{R}), > 1$

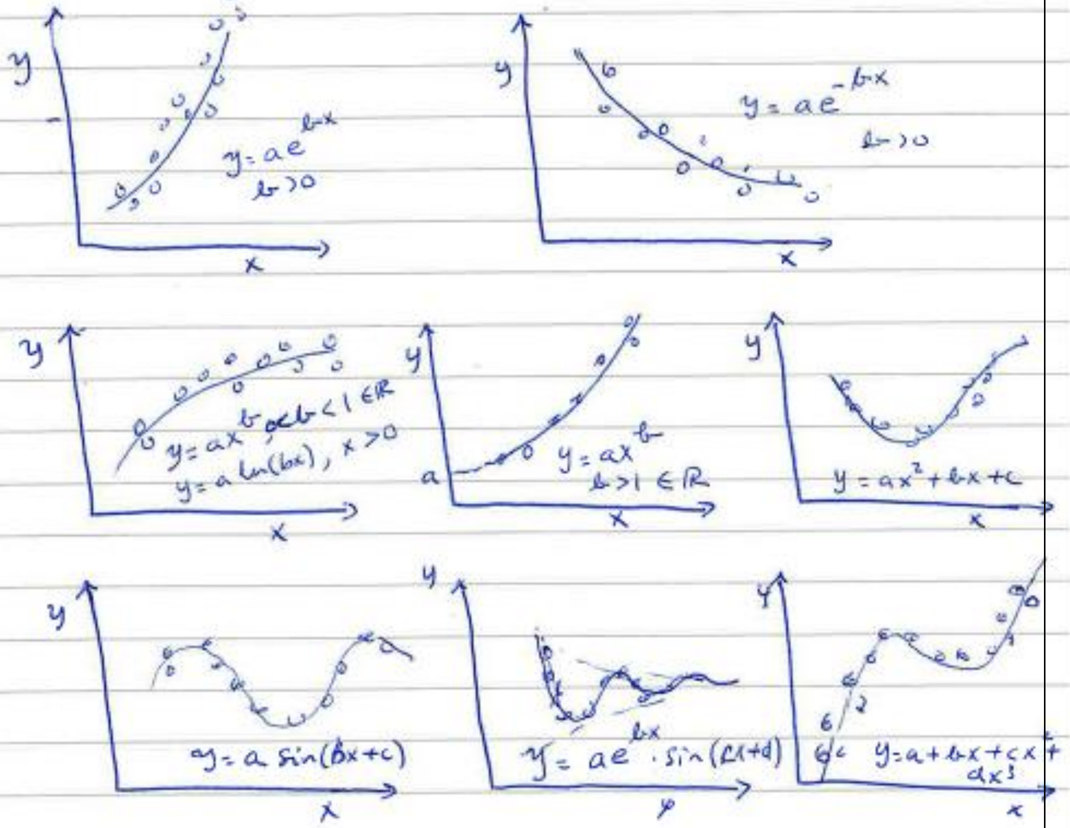
(c) Sin/cosine $y = a \sin(bx + c)$

(d) log/ln $y = a \ln(bx + c)$

(e) Combined fn $y = ae^{\beta x} \cdot \cos(bx + c)$

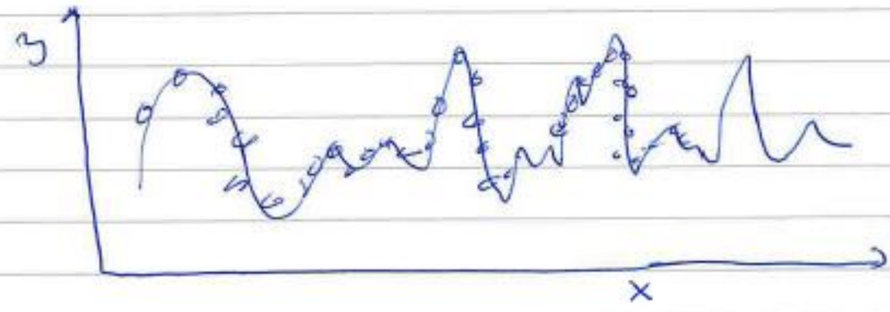
We select/estimate the above mapping function based on the trend of input/output data

Examples non-linear function



(III) If the dataset shows a complex trend and difficult to estimate the mapping function, we use deep learning (Artificial ~~intelligence~~ neural network) ML to build the model by training the dataset

Example

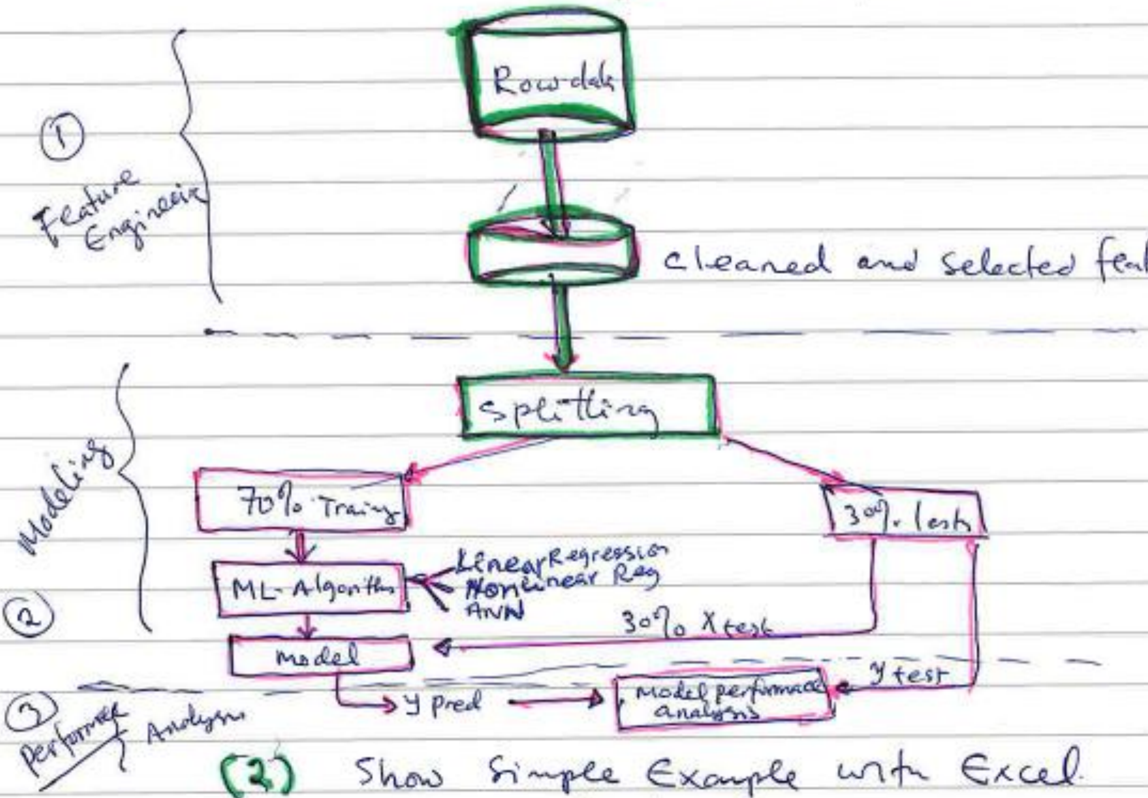


2.0 Machine Learning workflow

Introduction

Machine Learning modeling.
" "

(1) ML Workflow
- (1) Feature Engineering
- (2) modeling and prediction
- (3) Model performance evaluation



(2) Show Simple Example with Excel
→ Step by step
→ Feature Engineering
– Modeling
– Model performance analysis

(3) Adress Question about the blackbox
(1) How builtin library do modeling?
(2) How " " do performance analysis?

3.0 Source of Data

During the course, we will use two types of data

3.1 (a) Synthetic Data:

The data will be generated from physics based models

Examples

- Projectile motion → with/and without air/wind speed + drag
- Rectilinear motion
- Simple harmonic motion

3.2 (b) Field data

- The data is measured using Sensors.
- → It is a big data
- → Data is expected to contain
  - Noise
  - duplicates
  - Blanks / Un recorded
  - outliers + ...

Therefore the data must be pre-processed

Reason for using synthetic data is to VERIFY How ML model recover the known Physics model from which the data has been generated from.

4.0 Summary

We will learn data driven modelling both the concept and practical.

These are

a) • Linear Regression

b) • Non-Linear curvefitting

c) • ANN regression

> The three methods require input-output dataset.

> Before modelling, data must be PRE-PROCESSSED and the RIGHT FEATURES must be selected!

> Then, the input/output data will be trained – to develope 'ML Trained Model

> The trained ML-Model must be acessed its prediction accuracy

> Once we pass the model, we will use the model for application