Time Series 101: Learning from History

Lisa Ong, NUS-ISS

June 11, 2020

Speaker Introduction

iss.nus.edu.sg/about-us/staff/detail/203/Lisa%20ONG

linkedin.com/in/lisaong

github.com/lisaong

What I Teach

SOFTWARE SYSTEMS
NICF- Designing Intelligent Edge Computing (SF)

SOFTWARE SYSTEMS
NICF- Humanizing Smart Systems (SF)

STACKUP - STARTUP TECH TALENT DEVELOPMENT
NICF- Sequence Modeling with Deep Learning (SF)

STACKUP - STARTUP TECH TALENT DEVELOPMENT
NICF- Data and Feature Engineering for Machine Learning (SF)

STACKUP - STARTUP TECH TALENT DEVELOPMENT
NICF- Supervised and Unsupervised Modeling with Machine Learning (SF)

STACKUP - STARTUP TECH TALENT DEVELOPMENT
NICF- Feature Extraction and Supervised Modeling with Deep Learning (SF)

Topics

- **Definition** of Time Series
- **Applications** of Time Series data
- **Preparing** Time Series data
- **Training** Time Series data with Tensorflow-Keras
- **Next Steps**

Time Series Definition

Wikipedia:

A time series is a **series** of data points listed **in time order**.

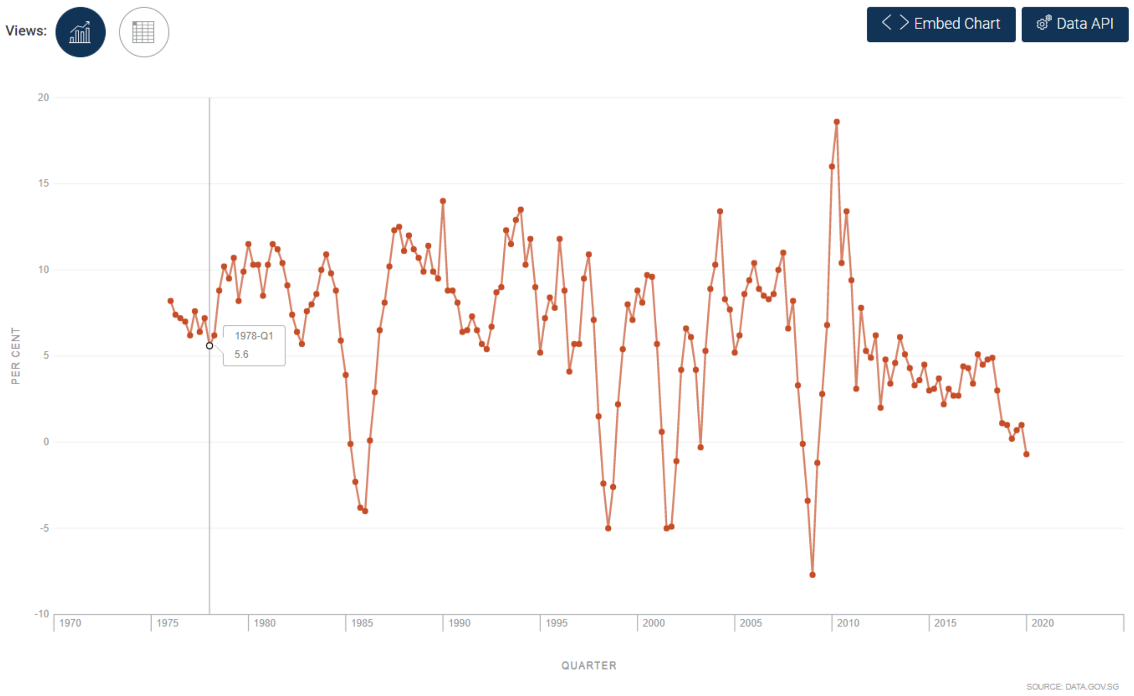
Most commonly, a **sequence** taken at **successive equally spaced points** in time.

Thus it is a sequence of **discrete-time data.**

You can plot the data with **time as the "x-axis".**

Time Sequential Equally-spaced Signal

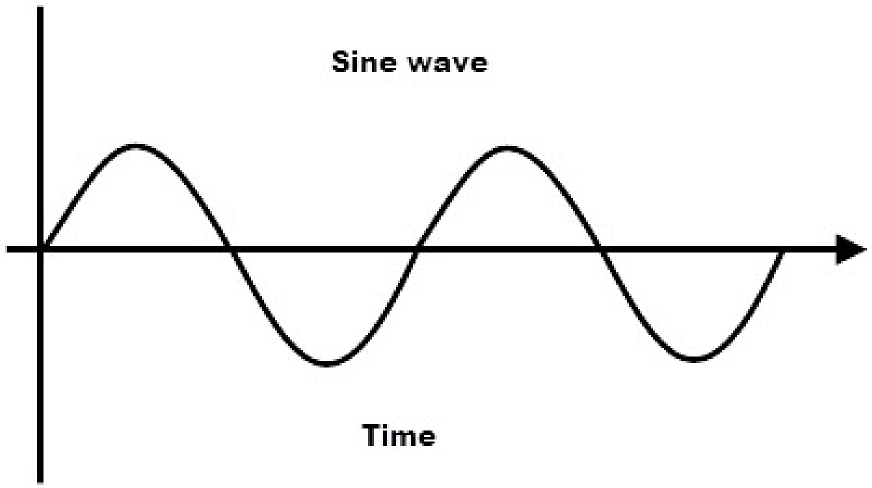
data.gov.sg/dataset/gross-domestic-product-in-chained-2015-dollars-year-on-year-growth-rate-quarterly

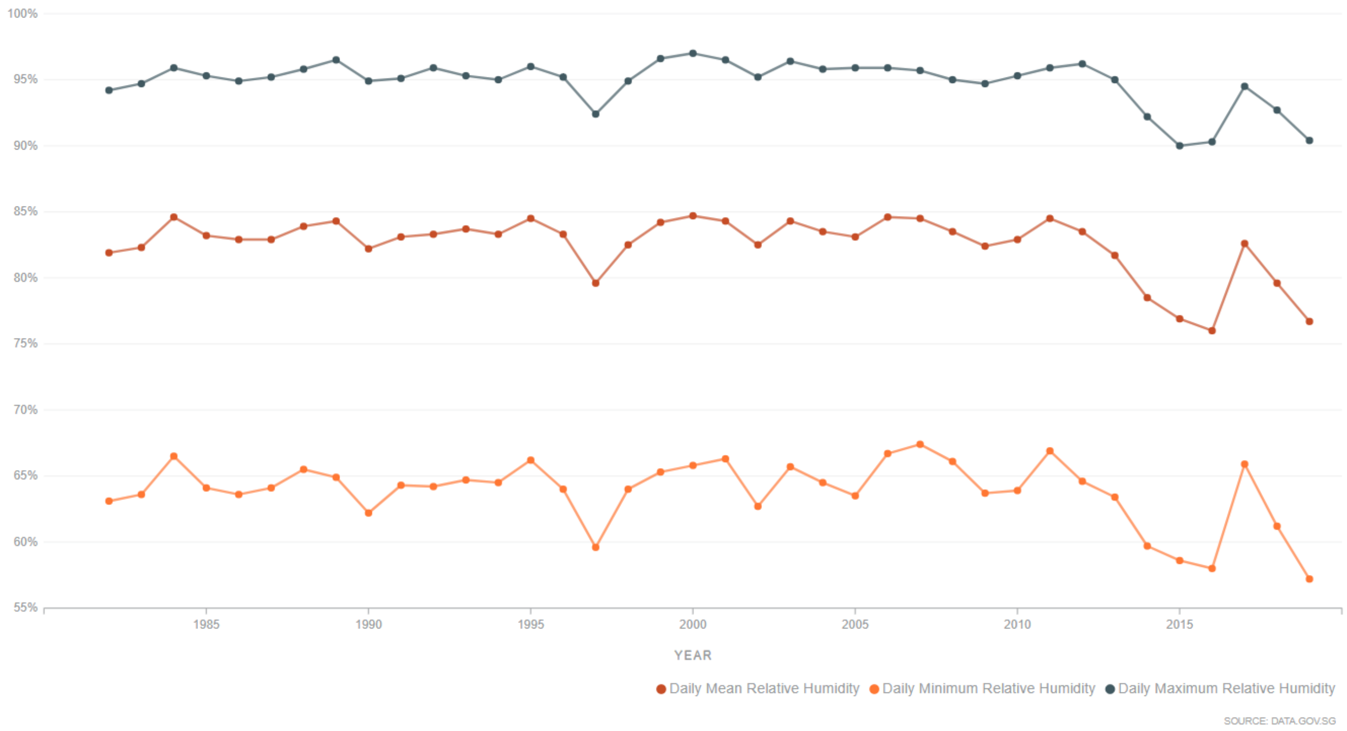


Quarter	Level 1	Value (Per Cent)
2020-Q1	GDP In Chained (2015) Dollars	-0.7
2020-Q1	GDP In Chained (2015) Dollars	-0.7
2019-Q4	GDP In Chained (2015) Dollars	1
2019-Q4	GDP In Chained (2015) Dollars	1
2019-Q3	GDP In Chained (2015) Dollars	0.7
2019-Q2	GDP In Chained (2015) Dollars	0.2
2019-Q1	GDP In Chained (2015) Dollars	1
2018-Q4	GDP In Chained (2015) Dollars	1.1
2018-Q3	GDP In Chained (2015) Dollars	3
2018-Q2	GDP In Chained (2015) Dollars	4.9
2018-Q1	GDP In Chained (2015) Dollars	4.8
2017-Q4	GDP In Chained (2015) Dollars	4.5
2017-Q3	GDP In Chained (2015) Dollars	5.1
2017-Q2	GDP In Chained (2015) Dollars	3.4

Other examples: Stock prices, Weather, Sales Revenues, Sensor Measurements

Can be Periodic, Cyclical, Seasonal, …





data.gov.sg/dataset/relative-humidity-annual-mean

What if I have dates in my dataset?

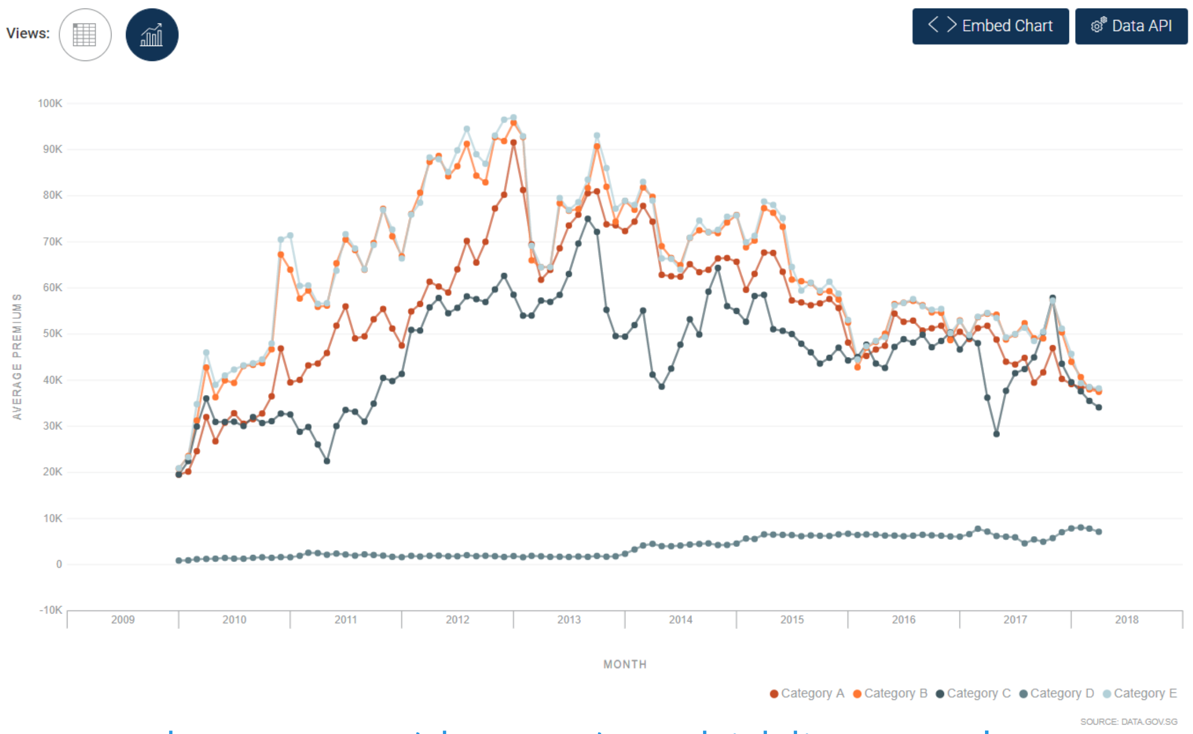
Purchase Order ID	Purchase Order Date	Ship Date	Shipped Via	Invoice Amount	Invoice Number	Inv Date	Terms	Paid
				Account Balance				
Customer: MBF Corp				**-$2,607.75**				
17 Orders				**$37,130.33**				
MBFID#1111	02/05/18	02/17/18	UPS Ground	$598.00	invoice #123	03/19/18	Net 30	✓
MBFID#1112	02/07/18	02/19/18	UPS Ground	$2,362.50	invoice #124	03/21/18	Net 30	✓
MBFID#1113	02/07/18	02/19/18	UPS 2nd Day Air	$1,569.50	invoice #125	03/21/18	Net 30	✓
MBFID#1114	02/08/18	02/20/18	Fedex	$1,097.07	invoice #126	03/22/18	2/10, Net 30	✓
MBFID#1115	02/09/18	02/21/18	UPS Ground	$252.45	invoice #127	03/23/18	Net 60	✓
MBFID#1116	02/09/18	02/21/18	UPS Ground	$2,164.96	invoice #128	03/23/18	Net 30	
MBFID#1117	02/09/18	02/21/18	UPS Ground	$851.54	invoice #129	03/23/18	Net 30	✓
MBFID#1118	02/12/18	02/24/18	UPS 2nd Day Air	$3,320.13	invoice #130	03/26/18	Net 60	✓
MBFID#1119	02/13/18	02/25/18	Fedex	$2,883.30	invoice #131	03/27/18	Net 60	✓
MBFID#1120	02/14/18	02/26/18	UPS Ground	$2,719.12	invoice #132	03/28/18	Net 60	✓
MBFID#1121	02/14/18	02/26/18	UPS 2nd Day Air	$4,337.37	invoice #133	03/28/18	2/10, Net 30	✓
MBFID#1122	02/14/18	02/26/18	Fedex	$719.00	invoice #134	03/28/18	2/10, Net 30	✓
MBFID#1123	02/19/18	03/03/18	UPS Ground	$3,703.51	invoice #135	04/02/18	Net 30	✓
MBFID#1124	02/20/18	03/04/18	Fedex	$2,274.65	invoice #136	04/03/18	Net 30	✓

Customer invoices

Q: Is there is a sequence of the <u>same</u> data point?
(E.g. are they purchases of the same item?)

What if I have multiple series?

Monthly COE Premiums by Category A, B, C, D, E

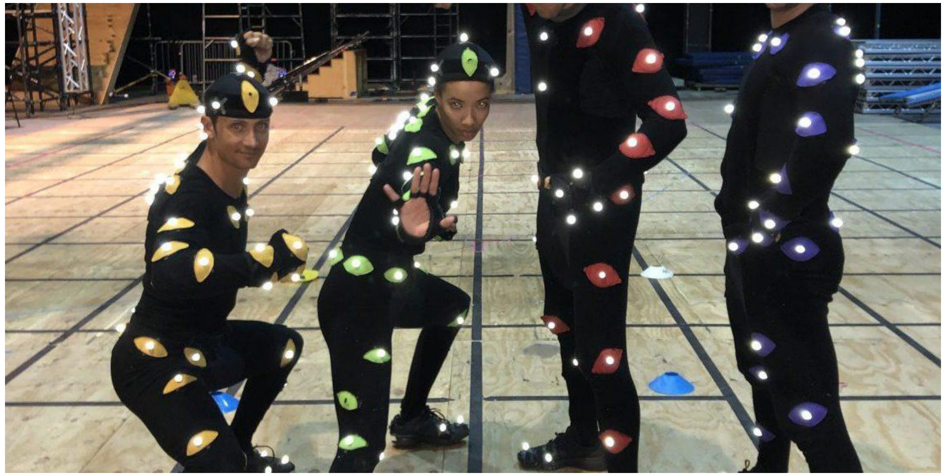


Month	Bidding No (No. of Bids)	Vehicle Class	Quota (No. of Quota)	Bids Success (No. of Successful Bids)	Bids Received (No. of Bids Recieved)	Premium (S$)
2020-03	1	Category A	978	973	1,436	32,699
2020-03	1	Category B	987	987	1,347	32,801
2020-03	1	Category C	315	315	504	24,202
2020-03	1	Category D	593	587	785	4,310
2020-03	1	Category E	333	331	512	32,500
2020-03	2	Category A	982	962	1,421	31,210
2020-03	2	Category B	992	943	1,366	30,012
2020-03	2	Category C	448	448	708	22,002
2020-03	2	Category D	581	576	757	4,489
2020-03	2	Category E	331	324	503	32,500
2020-03	1	Category A	978	973	1,436	32,699
2020-03	1	Category B	987	987	1,347	32,801
2020-03	1	Category C	315	315	504	24,202
2020-03	1	Category D	593	587	785	4,310
2020-03	1	Category E	333	331	512	32,500

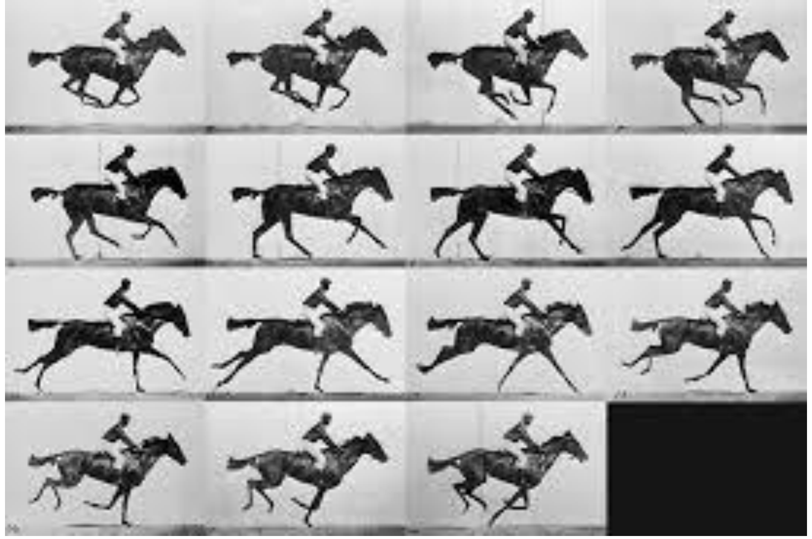
data.gov.sg/dataset/coe-bidding-results

Q: Are they independent or dependent?
(E.g. can Cat A be used to predict Cat E?)

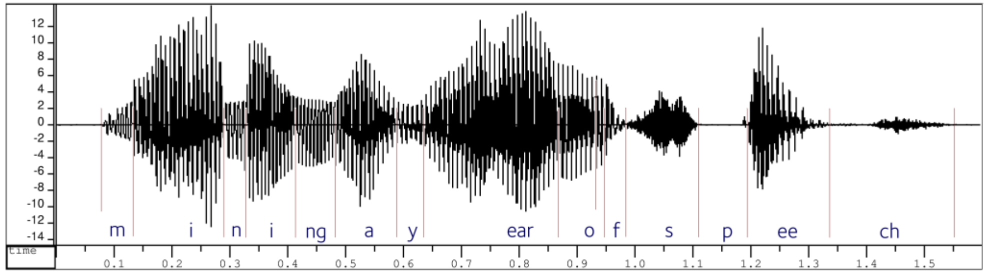
How about video, audio, motion?



www.chasearmitage.com/motion-captur/



nofilmschool.com

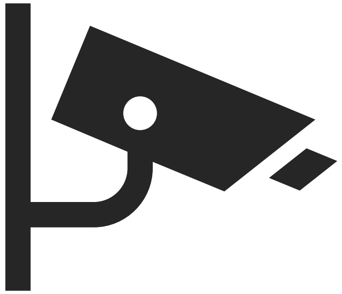


www.phon.ox.ac.uk

Specialised techniques:
e.g. Signal Processing, Recurrent + Convolutional Neural Networks, ...

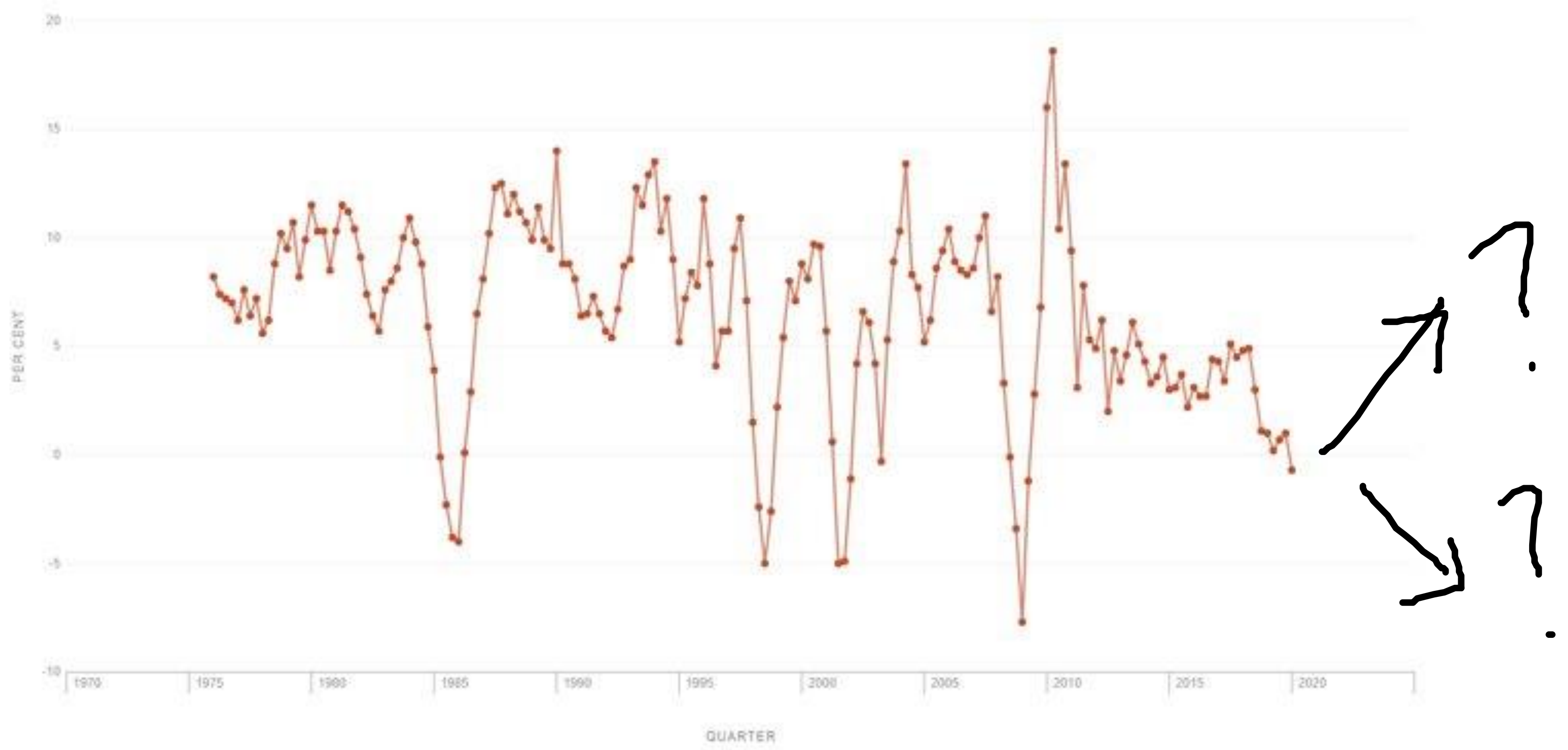
2 Common Applications

Predict the next value

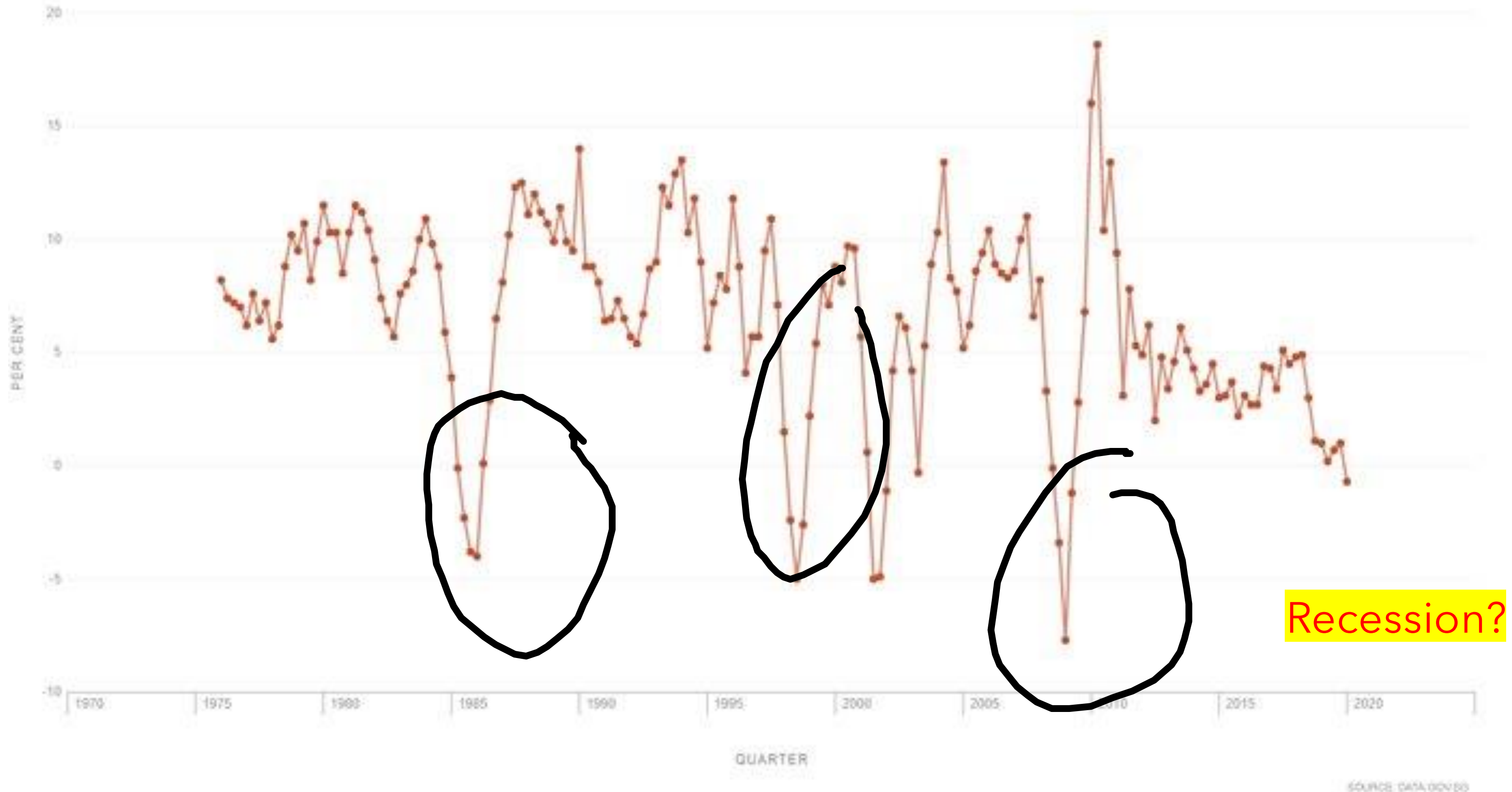


Recognise a pattern

Forecasting: Predict the next value(s)



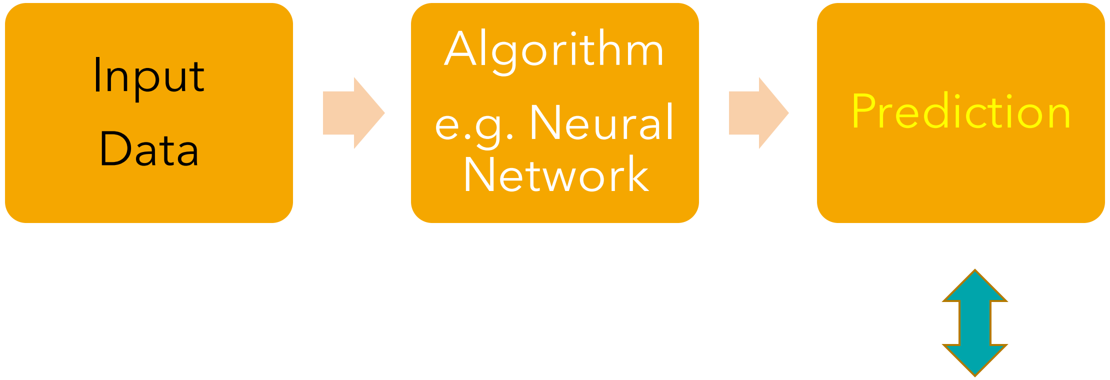
Pattern Recognition: Identify a Pattern



Applying Machine Learning or Deep Learning

A Machine Learning (ML) System is one where **input** data is passed into a trained **algorithm** to generate a **prediction. Deep learning** is a subset of Machine Learning.

To apply ML, we must first **formulate our problem** in terms of **inputs** and **targets.**



During "Training": Prediction is Compared with **Actual (Target) value** to **calculate the error** and **update the Algorithm**

How to formulate the ML problem?

- What are the inputs?
- What are the targets?

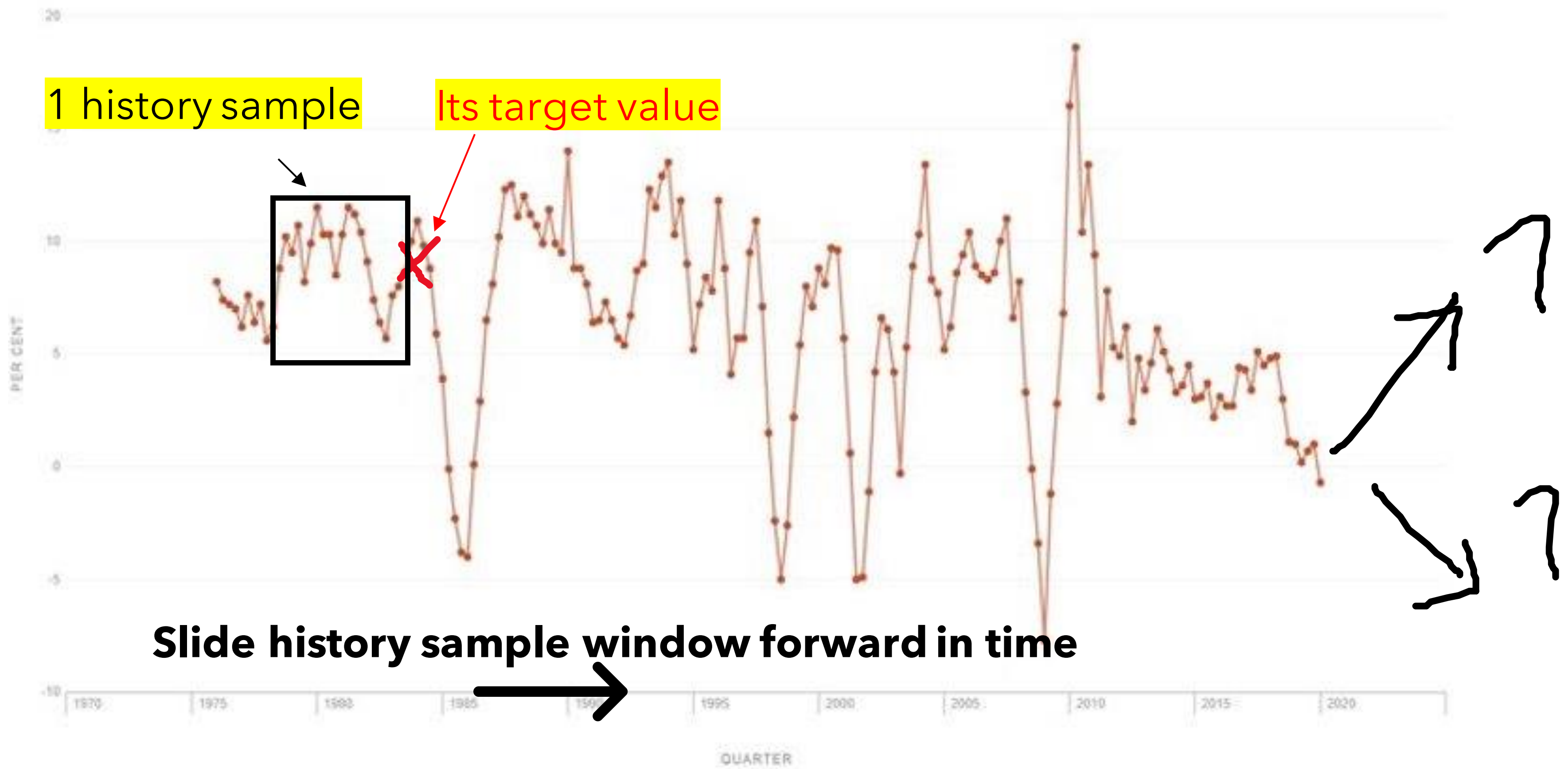
t	x
1	10
2	20
3	35
4	88
…	…

→

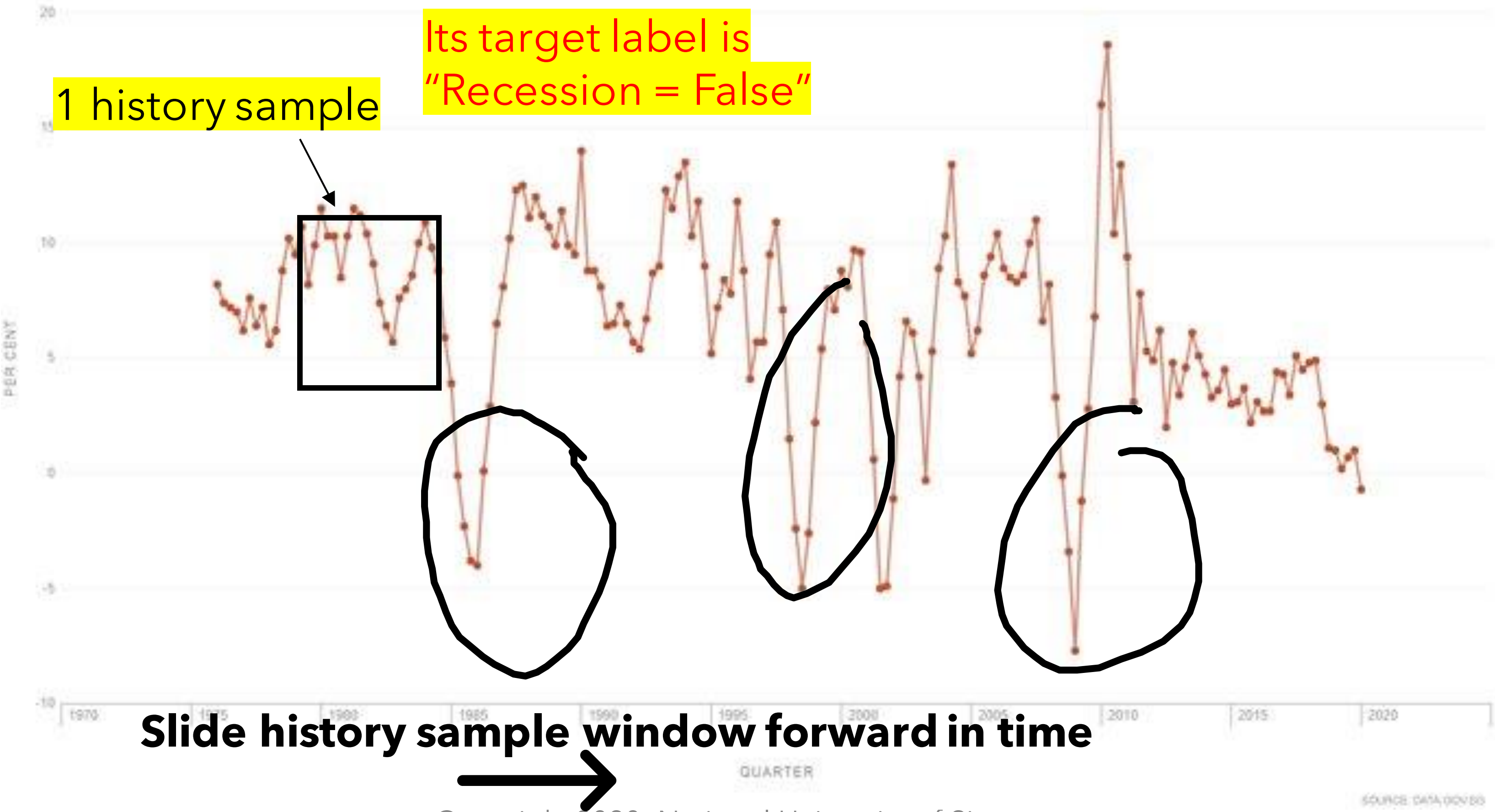
t	x [t]	x [t+1]	y
1	10	20	?
2	20	35	?
3	35	88	?

Each row is a **history (X) sample** and **its target (y)** at **time step t**. Target depends on the application.

Forecasting: History and Targets



Pattern Recognition: History and Targets



Why use a Window?

1. More than 1 value is needed to **determine direction**

2. Preserve the **chronological (sequence)-ordering** of the data, i.e. **"timeseries-ness"**

Given **only 1 value**, **will the next one go up or down???**

Quarter	Level 1	Value (Per Cent)
2020-Q1	GDP In Chained (2015) Dollars	-0.7

If **sequence order** is **not chronological,** **no longer a time series!**

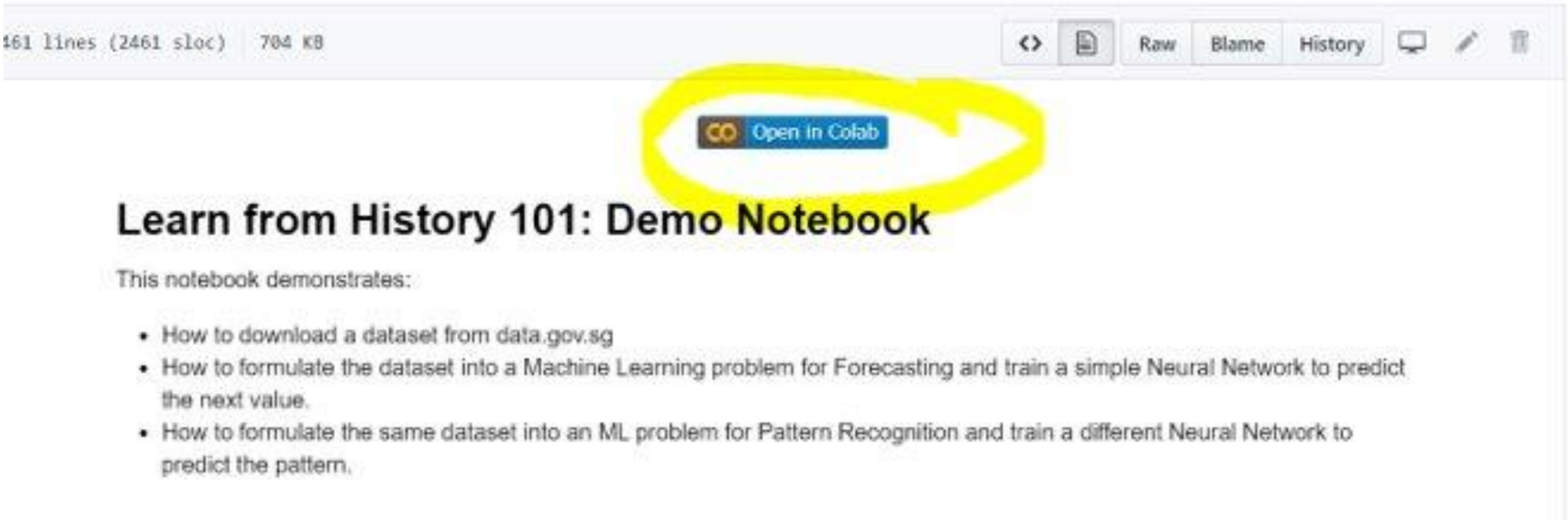
2020-Q1	GDP In Chained (2015) Dollars	-0.7
2019-Q4	GDP In Chained (2015) Dollars	1
2019-Q4	GDP In Chained (2015) Dollars	1
2019-Q3	GDP In Chained (2015) Dollars	0.7
2019-Q2	GDP In Chained (2015) Dollars	0.2
2019-Q1	GDP In Chained (2015) Dollars	1

Python Walkthrough

URL: bit.ly/iss-history101

- Loading a Time Series dataset from data.gov.sg
- Finding the "right" window size
- Creating a Rolling Window dataset
- Training a Neural Network for Forecasting

Google Colab Notebook: bit.ly/iss-history101

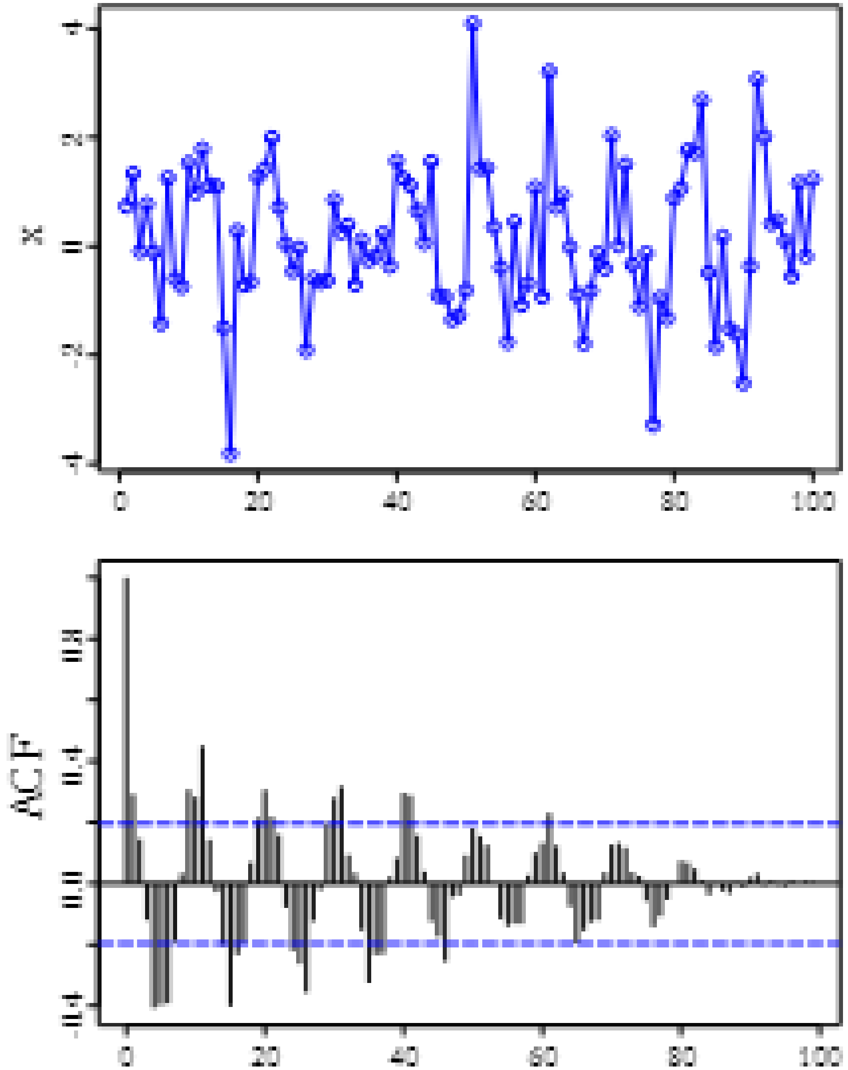


Finding the "Right" Window size

- Problem-specific
  - Forecasting: Auto-correlation or Partial Auto-correlation
  - Pattern Recognition: Observed length of the pattern
  - Both involve domain knowledge
- **Auto-correlation** is simplest, but only applies to **Forecasting**

Auto-correlation

- How dependent is the current value from historical values
    - Large (positive or negative) values: good
    - Near zero: random noise
- This is computed using a similar "**sliding window**" technique and measured at **different window sizes.**
- How to use: look for **auto-correlation exceeding the confidence band (95%)**

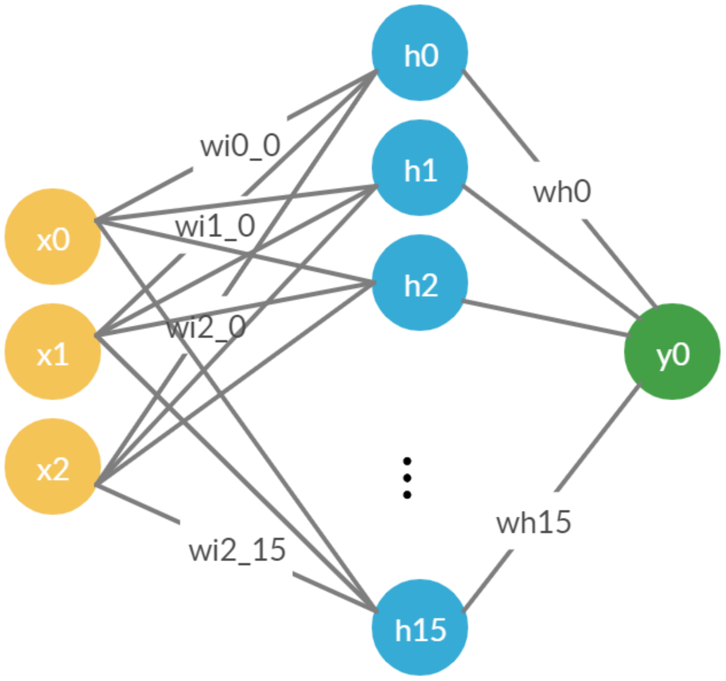


Wikipedia

Applying a Neural Network

- Simple Neural Network using **Tensorflow-Keras**
  - Multi-layer Perceptron (next slide)
- Check for **Overfitting**
  - When an algorithm only gets good predictions for the **training data**
- Get predictions

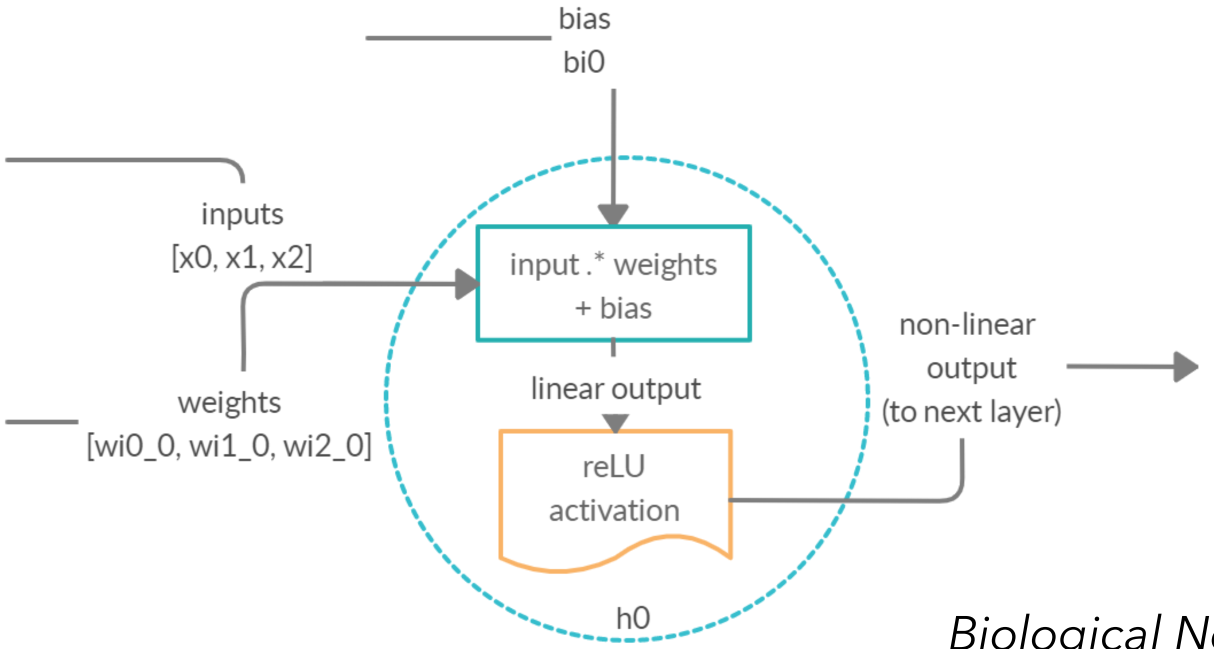
Multi-layer Perceptron (2 layers)



Parameters:

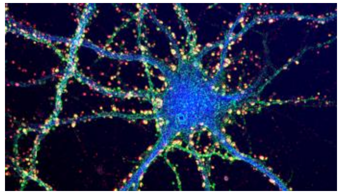
3x16 = 48 weights
16 biases

16x1=16 weights
1 bias



Neuron

Biological Neuron



brainfacts.org

Common Issues

- Overfitting: Neural Network doesn't do well with unseen data
- Predicted values are too smooth
- Bad data, incorrect labels

Python Walkthrough (continued)

- Reusing the same Time Series dataset
- Generating labels for ***Recession = True***
- Creating a Rolling Window dataset
- Training a Neural Network to detect ***Recession = True***

Summary

	Forecasting	Pattern Recognition
Target	Next GDP % value	True / False
History Window Size	Determined by the help of Auto-correlation	Determined by the pattern length
Output	Decimal values	Probabilities (using sigmoid activation)
Metrics	Mean Squared Error, Mean Absolute Error	Classification Accuracy
Training Loss Function	Mean Squared Error	Binary Cross Entropy Loss
Task	Regression	Classification

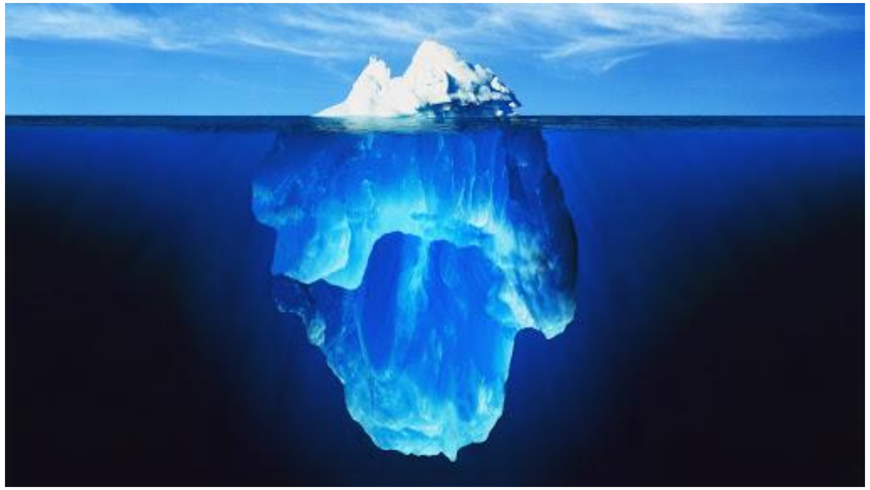
We've only scratched the surface

Other algorithms that can use "Windowing":

- Non-Neural Nets (e.g. Random Forest, Linear Regression)
- Recurrent Neural Nets, Convolutional Neural Nets

Statistical Methods that implicitly do "Windowing":

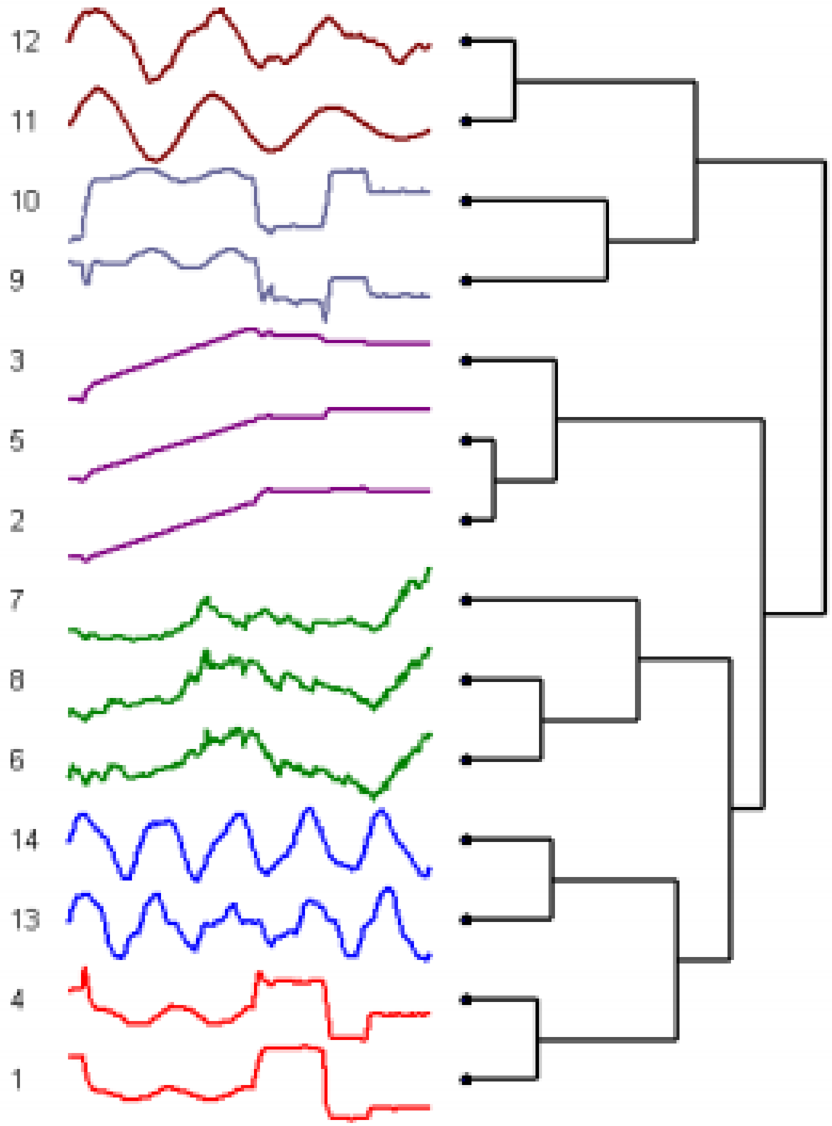
- AR, MA, ARIMA, SARIMA family: via the "p" and/or "q" parameter(s)
- www.statsmodels.org/stable/tsa.html



Articles by **Jason Brownlee:** machinelearningmastery.com/category/time-series/

More Applications of Time Series

- Clustering
  - Separating signal "patterns" into groups
  - Unsupervised or Semi-supervised learning
  - **No target label or value**
- Data generation
  - Test data, audio waveforms, etc
  - **A sequence of target values**



www.hanselsolutions.com

Upcoming Webinars and Courses

Webinar: Text Processing using the spaCy library

26 June

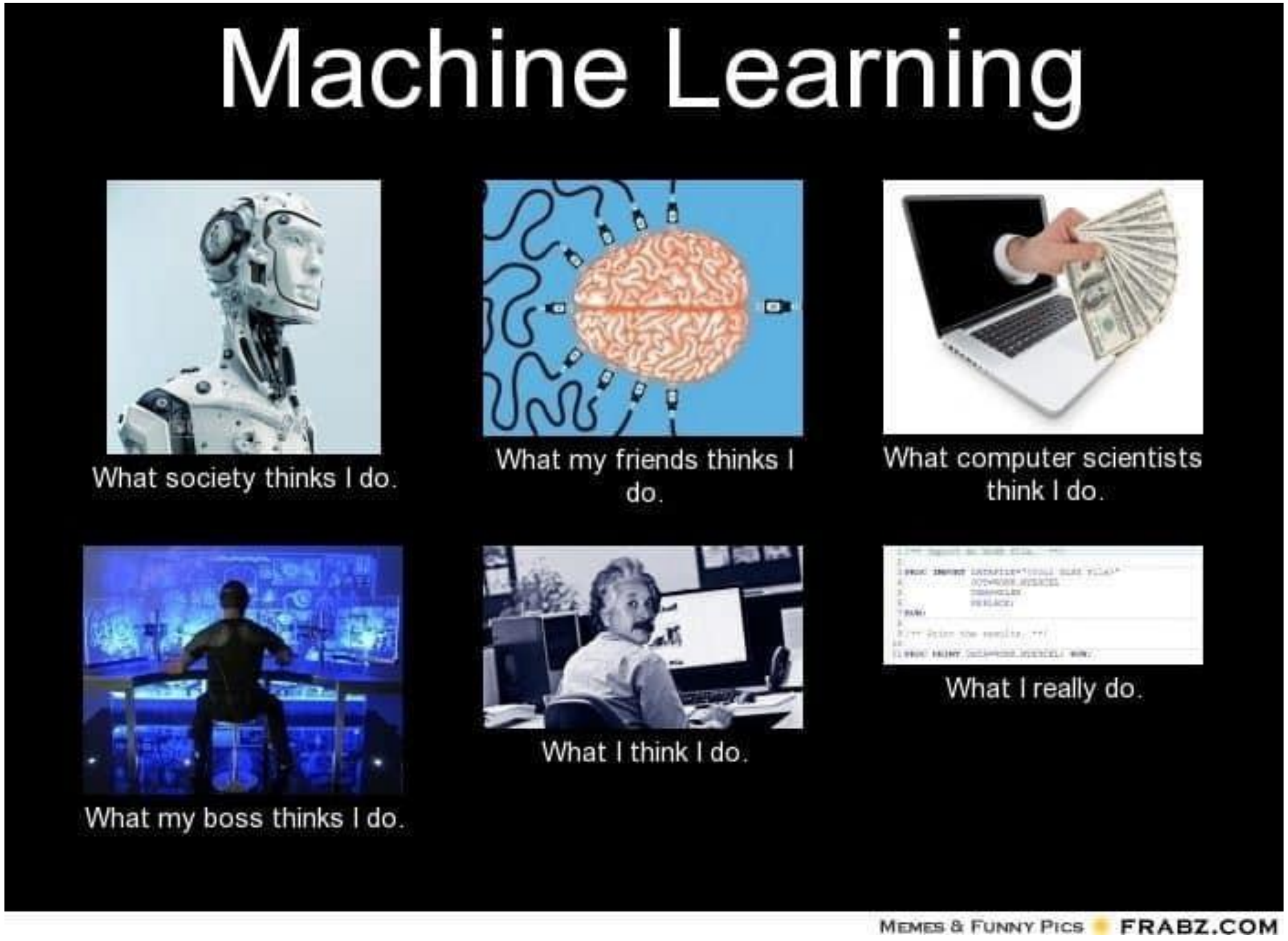
Short Course: Autonomous Decision-making with Reinfocement Learning

13 July

4 July

4-Course Series: Machine Learning-Driven Data Science: Part Time

Thank you!!


Machine Learning
What society thinks I do.
What my friends thinks I do.
What computer scientists think I do.
What my boss thinks I do.
What I think I do.
What I really do.
MEMES & FUNNY PICS FRABZ.COM

Feedback Survey

- Help us improve your experience
  - Scan the QR code or go to https://nus.edu/37jllHn
  - Presentation slides and recording will be shared after completion & successful submission of feedback
  - Allow 3-4 working days to receive the slides, upon successful submission.
  - Alternatively, write to us at issmarketing@nus.edu.sg