goML Phase 2

[Class Materials]

https://bit.ly/goML_Jan2021

Good morning! Welcome to phase 2 of the programme.

We will start at 9am sharp or when majority of participants are online.

Meanwhile, grab your coffee or tea, sit back and relax.

PART2

Mr Seow Khee Wei / Mr Shubham Khare / Mr Koay Seng Tian

Programme Overview

goML Phase 1

Activity 1.5 – Data Preparation

- 3 cleaning exercises
  - non-numeric characters are set to null
  - non-numeric characters are dropped
  - Splitting the data for training and testing.

Step 1:
Watch and listen to the instructor's demonstration

15 mins

Target to finish by 16:58

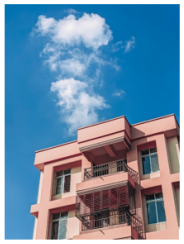
Step 2:
Work through the activities

30 mins

Individual Activity

Activity 2.1

- Training a Regression Model using Linear Regression



sq feet	num bedrooms	num bathrooms	sale price
785	2	2	170461
1477	2	2	271651
712	1	1	139912

Bedrooms	Bathrooms	Sq. Feet	Sale Price
3	2	2000	???

Exercise:
- Predict how much a 1500 sqft, 3 bedrooms with 3 bathrooms cost.

Keras pandas learn TensorFlow

Step 1:
Watch and listen to the instructor's demonstration

15 mins

Step 2:
Work through the activities

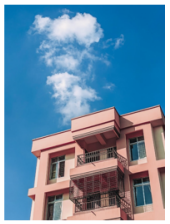
Target to finish by 10:10

15 mins

Individual Activity

Activity 2.2

Use tensorflow & keras to train and use a neural network model to estimate the price of a house



sq feet	num bedrooms	num bathrooms	sale price
785	2	2	170461
1477	2	2	271651
712	1	1	139912

Bedrooms	Bathrooms	Sq. Feet	Sale Price
3	2	2000	???

Exercise:
- Add a hidden layer with 100 nodes. Did the performance improve?

Step 1:
Watch and listen to the instructor's demonstration

20 mins

Step 2:
Work through the activities

Target to finish by 11:20

10 mins

Individual Activity

Activity 2.3 - HDB price predictor

Exercises:
- Add different regression algorithms and evaluate performances

Target to finish by 11:50

Step 1:
Watch and listen to the instructor's demonstration

15 mins

Step 2:
Work through the activities

15 mins

Individual Activity

Activity 2.4 – Classification with Logistic Regression

ATTRIBUTE	DESCRIPTION	VALUE
Preg	Number of pregnancies	[0 – 17]
Plas	Plasma glucose concentration in an oral glucose tolerance test	[0-199]
Pres	Diastolic blood pressure	[0-122]
Skin	Triceps skin fold thickness	[0-99]
Insu	2-Hour serum insulin	[0-846]
Mass	Body mass index	[0-67]
Pedi	Diabetes pedigree function	[0-2.45]
Age	Age of an individual	[21-81]
class	Tested positive / negative	(0,1)

Exercises:
Is a patient with the following data diabetic?
preg:7, plas: 132, pres: 80,
skin: 30, test: 0, mass: 45.5, pedi: 0.547, age: 45

Target to finish by 13:50

learn

Step 1:
Watch and listen to the instructor's demonstration

15 mins

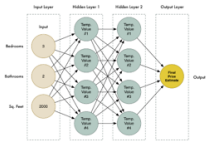
Step 2:
Work through the activities

15 mins

Individual Activity

Activity 2.5 – Classification with Neural Network

ATTRIBUTE	DESCRIPTION	VALUE
Preg	Number of pregnancies	[0 – 17]
Plas	Plasma glucose concentration in an oral glucose tolerance test	[0-199]
Pres	Diastolic blood pressure	[0-122]
Skin	Triceps skin fold thickness	[0-99]
Insu	2-Hour serum insulin	[0-846]
Mass	Body mass index	[0-67]
Pedi	Diabetes pedigree function	[0-2.45]
Age	Age of an individual	[21-81]
class	Tested positive / negative	(0,1)



Exercises:
Modify the hidden layers with additional nodes. Did the performance improve?

Target to finish by 14:15

Step 1:
Watch and listen to the instructor's demonstration

15 mins

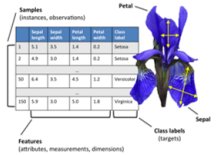
Step 2:
Work through the activities

15 mins

Individual Activity

goML Phase 1

Activity 2.6 – Multi Class Classification



Exercises:
Modify the hidden layers with additional nodes.
Did the performance improve?

Keras pandas learn TensorFlow

Step 1:
Watch and listen to the instructor's demonstration

10 mins

Target to finish by X:XX
Step 2:
- Do on your own

30 mins

Individual Activity

64

Activity 2.7 – Classifier Use Case

Dataset:
https://www.kaggle.com/c/titanic/data

Exercises:
Use a different classifier and compare the results

Keras pandas learn TensorFlow

Step 1:
Watch and listen to the instructor's demonstration

10 mins

Target to finish by 14:55
Step 2:
- Do on your own

30 mins

Individual Activity

65

Activity 2.8 – Hyperparameter tuning for Linear Regression

sq feet	num bedrooms	num bathrooms	sale price
785	2	2	170461
1477	2	2	271651
712	1	1	139912

Bedrooms	Bathrooms	Sq. Feet	Sale Price
3	2	2000	???

Exercise:

Step 1:
Watch and listen to the instructor's demonstration

15 mins

Step 2:
Work through the activities

Individual Activity

15 mins

78

Activity 2.9 – Hyperparameter tuning for Classifier

ATTRIBUTE	DESCRIPTION	VALUE
Preg	Number of pregnancies	[0 – 17]
Plas	Plasma glucose concentration in an oral glucose tolerance test	[0-199]
Pres	Diastolic blood pressure	[0-122]
Skin	Triceps skin fold thickness	[0-99]
Insu	2-Hour serum insulin	[0-846]
Mass	Body mass index	[0-67]
Pedi	Diabetes pedigree function	[0-2.45]
Age	Age of an individual	[21-81]
class	Tested positive / negative	(0,1)

learn

Step 1:
Watch and listen to the instructor's demonstration

15 mins

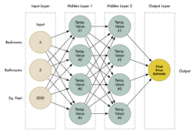
Step 2:
Work through the activities

Optional

15 mins

79

Activity 2.10 – Hyperparameter tuning for Neural Network



Target to finish by 15:45

Step 1:
Watch and listen to the instructor's demonstration

15 mins

Step 2:
Work through the activities

Optional

15 mins

80

goML Phase 2

Phase 2 (2 days) focuses on project development facilitation. Participants will form project groups of 1 to 4 participants with a real project in-hand. And a mentor will be assigned to facilitate the engagement of the real project amongst project participants.

Team Formation

- 2 to a team
- Domain:
  - Structured Data (Regression or Classification)

- To complete proposal template before phase 2 (24-25 Jan 2022)
  - Problem statement
  - Dataset

Programmes

Day 1	Project sharing with trainer Refining project scope Evaluation / Acceptance Criteria Dataset(s) - Data Understanding	Data Preparation - Data Labeling - Data Cleaning - Data Augmentation - Anonymization/De-nosing/Standardization/Normalisation
Day 2	Model Training & Evaluation - Baselining - Algorithm Selection	Hyperparameter tuning **Project Sharing (4-5 slides, 15 mins per team)**

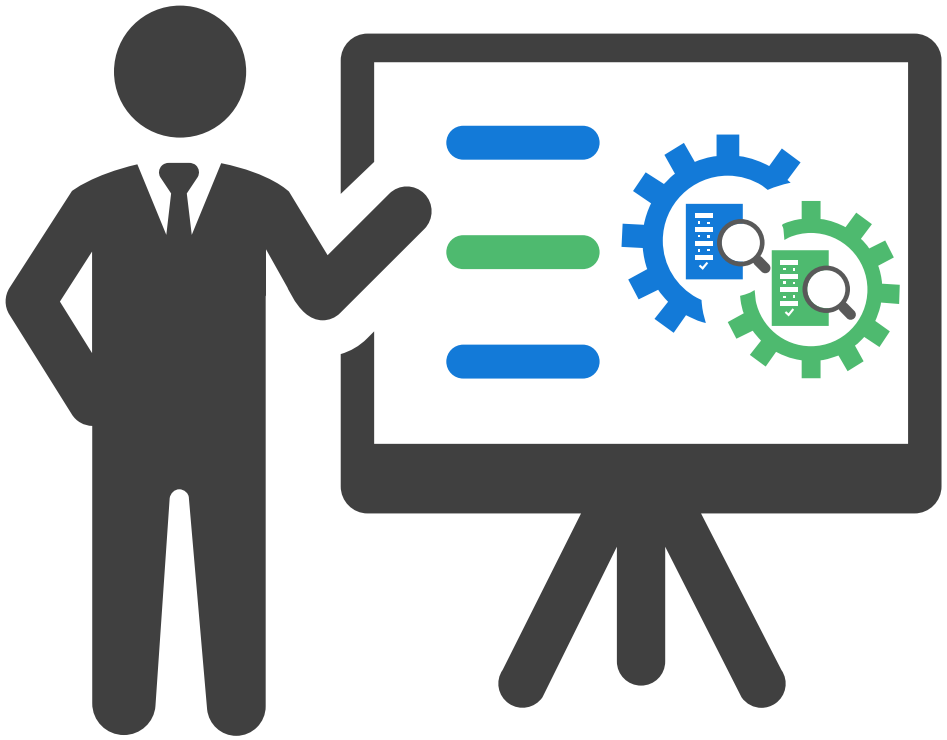
Team

Team	Member	Title	Trainer
Team 1	Hiroshi Fujiwara Lai Joo Qee		
Team 2	Goh Quan En Wong Ngan Seng Ezekiel	Arc scratch prediction using tool traces	
Team 3	Kyawt Ngone Thet Lee Zi Jia Timothy	KNOWLEDGE GRAPH on SMART SCADA system	
Team 4	Gurusubramanian Karthikeyan Johanes Eric Tan		
Team 5	Lee Chee Wai Gan Yin Ze	Toolcarts Image Recognition	
Team 6	Benedict Eli Abraham Shuek Lee Jin Yao		
Team 7	Teng Yi Fang Cheryl Oon Jianxiong	Machine alarms prediction	
Team 8	Liu Fuming Tan Yuen Min	GUI Automation	

Online Support

To provide details

Q&A



Email
seow_khee_wei@rp.edu.sg

Telegram
@kwseow

Source code:

Thank you