CE / CZ 4073

Data Science for Business
Semester 2 | 2017-2018

Instructor	Email	Office
Sourav SEN GUPTA	sg.sourav@ntu.edu.sg	N4-02b-66 | 6790 4587

Structure

20 Regular Lectures (one hour each)	8:30 AM to 10:30 AM on Fridays	Week 01 to 13 (minus holidays)
+4 Self-Learn Videos (one hour each)	At your own time and pace ☺	Weeks with Holiday (05 and 11)
+9 Regular Tutorials (one hour each)	2:30 PM to 3:30 PM on Fridays	Week 02 to 13 (minus holidays)
+2 Lectures (one hour) for Revision	8:30 AM to 10:30 AM on Fridays	Week 14 (only on 20th of April)
+2 Tutorials (one hour) for extra help	2:30 PM to 3:30 PM on Fridays	Week 01 and 14

No Laboratory is assigned for the course. You may feel free to bring your laptop to lectures if you want to explore the worksheets related to the recitation. You are also encouraged to bring your laptop to the tutorials, as most of the tutorials will be hands-on sessions. Office hours for Sourav are 10:00 AM to 12:00 Noon on Tuesdays, or by appointment over email.

Evaluation

Written Examination (40%) at the end of the Semester + Three Assignments (20% x 3 = 60%) during the Semester

The written examination will be conducted similar to other SCSE end-semester written examinations. The assignments will be data science related computing assignments in R. In total, four assignments will be posted during the semester, two before the recess week and two after, and the best three (out of the four) will be counted towards the final grade.

The assignments are all take-home assignments, to be submitted via email within the stipulated deadline (10 to 12 days).

Required Material

There is no single textbook for the course. The students are expected to be mature enough to follow the lectures and refer to multiple resources (mostly online), as and when required. The only mandatory component of the course is to learn R as the computing framework. It may be expected that the students will install R and R-Studio on their computers.

References and Resources

An Introduction to Statistical Learning (http://www-bcf.usc.edu/~gareth/ISL/) : James, Witten, Hastie, Tibshirani

Data Science for Business (http://data-science-for-biz.com/DSB/Home.html) : Provost and Fawcett

R Package (https://www.r-project.org/) and **RStudio** (https://www.rstudio.com/) : Download and Install

Additional resources, if required, will be shared with the students from time and again, in the Lectures and/or Tutorials. Almost all of these resources will either be online (free and open source) books or online (freely available) lecture videos.

Course Schedule

Week	Lec / Tut	Topic	Remarks
01 (19/1)	Lecture 01	Motivation and Introduction – What is Data Science?	Slides – No worksheet
	Lecture 02	Basic concepts of Statistics and Data Handling in R	Worksheet in R
	Tutorial 00	*Installation of R and R-Studio*	*Optional session*
02 (26/1)	Lecture 03	Prediction in Business – Introduction to Linear Models	Slides with Worksheet
	Lecture 04	Linear Regression – Training, Minimization and Variance	Worksheet in R
	Tutorial 01	*Introduction to R and Gradient Descent for Regression*	*Hands-on Demo + Theory*
03 (02/2)	Lecture 05	Linear Regression – Parameter Estimation and Inference	Slides with Worksheet
	Lecture 06	Linear Regression – Model fitting in Multi-Variate Setting	Worksheet in R
	Tutorial 02	*Linear Regression - Training, Minimization and Prediction*	*Q&A + Hands-on Demo*
04 (09/2)	Lecture 07	Classification in Business – Motivation for Decision Trees	Slides with Worksheet
	Lecture 08	Aggregating Models – Classification using Random Forests	Worksheet in R
	Tutorial 03	*Linear Regression - Complete Model Building Exercise*	*Q&A + Hands-on Demo*
05 (16/2)	Self-Learn 01	Classification using Linear Models – Logistic Regression	Video + Worksheet (holiday)
	Self-Learn 02	Performance of a Classification Model – Accuracy, ROC	Video + Worksheet (holiday)
06 (23/2)	Lecture 09	Choosing a Model – How to avoid Overfitting?	Slides with Worksheet
	Lecture 10	Bias-Variance Trade-off and Cross-Validation	Worksheet in R
	Tutorial 04	*Trees and Forests + Performance of Classification Models*	*Q&A + Hands-on Demo*
07 (02/3)	Lecture 11	Regression using Tree Models and Gradient Boosting	e-Learning (recorded lecture)
	Lecture 12	Regression and Classification using Nearest Neighbors	e-Learning (recorded lecture)
	Tutorial 05	*Choosing a Model - Training, (Cross-)Validation and Test*	*e-Learning (sample code in R)*

Week	Lec / Tut	Topic	Remarks
08 (09/3)	None	None	Recess Week
09 (16/3)	Lecture 13	Notion of Distance and concept of Nearest Neighbors	Slides with Worksheet
	Lecture 14	Neighborhoods to Clusters – k-Means and Dendograms	Worksheet in R
	Tutorial 06	*Regression and Classification using Nearest Neighbors*	*Q&A + Hands-on Demo*
10 (23/3)	Lecture 15	Forecasting in Business – Fundamentals of Time Series	Slides with Worksheet
	Lecture 16	Time Series – Trends, Seasonality and Cycles in Data	Worksheet in R
	Tutorial 07	*Nearest Neighbors and Clustering (k-Means, Dendograms)*	*Q&A + Hands-on Demo*
11 (30/3)	Self-Learn 03	Detecting Anomalies – Outliers and Misfits in Clustering	Video + Worksheet (holiday)
	Self-Learn 04	Anomaly Detection in Multi-Variate Data and Time Series	Video + Worksheet (holiday)
12 (06/4)	Lecture 17	Visualizing Data through Low-Dimensional Embedding	Slides with Worksheet
	Lecture 18	Storyboarding – Dashboards and Visualizations in R	Worksheet in R
	Tutorial 08	*Time Series Analysis for Business Forecasting Problems*	*Q&A + Hands-on Demo*
13 (13/4)	Lecture 19	Data Analytic Thinking – Practical examples from Business	Invited Lectures (tentative)
	Lecture 20	Data Analytic Thinking – Practical examples from Business	Invited Lectures (tentative)
	Tutorial 09	*Dashboards and Visualizations in RShiny and Python*	*Hands-on session in R*
14 (20/4)	Lecture 21	Revision of all concepts discussed during the course	Slides with Worksheet
	Lecture 22	Revision of all concepts discussed during the course	Slides with Worksheet
	Tutorial 10	*Revision of all concepts discussed during the course*	*Optional session*

Assignment Schedule (tentative)

Assignment 1	Post : 5 February	Deadline : 19 February	Topic : Weeks 01 to 04
Assignment 2	Post : 23 February	Deadline : 5 March	Topic : Weeks 04 to 06
Assignment 3	Post : 16 March	Deadline : 26 March	Topic : Weeks 07 to 09
Assignment 4	Post : 6 April	Deadline : 16 April	Topic : Weeks 10 to 12