Course Outline

Engineering Data Analytics – ENGN8535

2021S1 version-1.0

Overview:

This is a graduate level (8000 series) course offered by the ANU College of Engineering and Computer Science, for Masters and PhD students majoring in Advanced Computing, or Mechatronics, Machine Learning, Computer Vision, Artificial Intelligence, Intelligent Systems, System Engineering, or Electronic Engineering.

Data-driven decision-making is an essential component of emerging engineering and computing systems that generate and consume very large amounts of sensing data from autonomous vehicles to digital pathology.

This course covers fundamental theories and methods in Data Science necessary for inferring useful information and identifying underlying patterns from often raw, incomplete, high dimensional, noisy and corrupted data that is present in real-life applications. It will provide students the opportunity to explore advanced solutions of data analytics such as dimensionality reduction, manifold learning, data clustering, sparse coding, dictionary learning, compressive sensing, nonlinear filtering, and generative data modelling.

Expected Learning Outcomes:

Upon successful completion, students will have the knowledge and skills to:

1. Describe a number of models for inference from data
2. Assess the strength and weakness of each of these models
3. Interpret the mathematical equations from Linear Algebra, Statistics, Probability Theory and numerical Optimisation used in the learning models
4. Implement efficient learning algorithms on a computer
5. Design test procedures in order to evaluate a model
6. Combine several models in order to gain better results

Assumed Knowledge

Linear algebra, Matrix Computation.
Probability & Statistics.
Computer Programming and Python Language.

2021 S1 Lecturers: Prof Hongdong Li (Convener) :

Professor's Office hours (Thursday morning 11:00am--12:00 noon, via ZOOM link TBA)

2021 TA/Tutors: Mr. Shidong Pan;
Mr. Vincent Cheng :
Tutors' office Hours : (Friday afternoon 4:00pm--5:00pm, via ZOOM link).

Related Fields:

Statistics and Data Science, Big-Data Analytics

Machine Learning and Pattern Recognition

Artificial Intelligence and Deep Learning

Signal Processing

Prescribed textbook:
None. Reading materials will be recommended.

Indicative assessment components (subject to change):

1. Homework x4 = 10%x4 = 40%
2. Computer Lab x4 = 7.5%x4 = 30%.
3. Term project x1 = 30%.

Syllabus:

Topic-1: Introduction

1.1 Python coding basics and Github/GitLab

Topic-2: Mathematics review
2.1 Linear Algebra and Matrix computation
2.2 Probability theory
2.3 Numerical Optimisation

Topic-3: Ranking and Recommendation algorithms

3.1 PageRank ranking. (eigen)
3.2 Recommendation system (SVD, NMF)

Topic-4: High dimensional data geometry
4.1 High dimensional Spaces
4.2 High dimensional Gaussians

Topic-5: Dimensionality Reduction

5.1 PCA
5.2 MDS
5.3 tSNE
5.4 AutoEncoder

Topic-6: Manifold Learning
6.1 IsoMap
6.2. LLE
6.3 Hessian Embedding (optional)
6.4 Laplace Embedding (optional)

Topic-7: Unsupervised Clustering

7.1. K-Means
7.2. GMM and EM
7.3. Spectral Clustering

Topic-8: Signal unmixing and BSS
8.1 BSS problem
8.2 ICA algorithm

Topic-9: Compressive Sensing

9.1. Under-constrained linear equations.
9.2. Sparse recovery
9.3 Dictionary Learning

==== ===

- **Weekly Study-load = 10 hours ~ 11 hours per week.**

About 2~3 hours per week on online lecture/tutorial.
About 8 hours per week on lecture-note reading, homework and computer labs, and on term project.

Weekly Study Plan (ver 2.0)

Week/Task	Thursday Lecture	Friday Lecture/Tute	Homework and Computer Lab	Term Project (Team project of 2~3 students per team.)	Note
week-1: (22nd Feb 2021 --)	Introduction; Math-1 (lin alg)	Tute-1: Python/GitLab; And Homework submission instruction;			
2 (1st Mar --)	Math-2 (probability)	Tute-2: Math-3 (unconstrained and constrained optimisation)	Hw1 /CLab-1 announced	Tentatively form your project team (2~3 students per team)	
3 (8th Mar--)	Ranking (PageRank) Recommendation (collaborative filter)	Recommendation (cont.)			
4 (15	High dimensional data	Tute-3: HW-1 tutorial ;	Hw1/CLab-1 due on Thursday	Project topics announced	

March/ICCV)		**Project topics announced**	evening;		
5	Dimension Reduction (PCA, MDS, AutoEncoder)	Data visualization (tSNE); Project topic Q&A.	Hw2/Clab-2 announced	Form your project team.	
6 (Good Friday)	Manifold Learning: IsoMap + LLE	Good Friday holiday; Project team forming and proposal due;		Project team forming. Project proposal due (one page A4) containing the following information: Team name, List of your team members (names+ uIDs), project topic, content, project timeline, tentative workload distribution.	
B1 (teaching break, 5-April --)			Hw2/Clab-2 due on Thursday C.O.B.	Start working on your term-project.	
B2 (teaching break, 12 April --)					
7 (19 April --)	Data clustering (Keans, GMM)	Data clustering (spectral clustering)	Hw3/Clab-3 announced		
8	BSS and ICA	Tute-4: Hw2 tutorial ; Term project check-point and Q&A .			
9	Compressive Sensing -1	Tute-5: Hw3 tutorial	Hw3/Clab-3 due on Thursday		

			cob. Hw4/Clab-4 announced		
10	CS-2	Tute-6: Term project check-point and Q&A .			
11	CS-3		Hw4/Clab-4 due.		
12 (24 May -- 28 May.)	TBC (flexible)	TBC (flexible)			
13 (31 st May --)	Term project report due: on 2nd of June.			Term Project report due 2nd of June	
14 Exam Weeks					