QUEEN'S UNIVERSITY BELFAST
SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Data Analysis & Visualisation

CSC3062

BEng (CS & SE), MEng (CS & SE), BIT & CIT

Dr Reza Rafiee

23rd September 2019

About me

- Expert in Data Analytics and Machine Learning
  - BEng (in computer science), MSc (in AI & Robotics) and PhD (in machine learning) & more than 5 years post-PhD research experience in data analytics, machine learning, bioinformatics and software development
- Passionate about machine learning based software development (specifically medical applications)

https://pure.qub.ac.uk/portal/en/persons/reza-rafiee(d0fb76ef-bc76-4a12-ba27-476be1a46cd1).html

Be Focused, passionate and driven

This course aims to …

- Explain how data analytics is used in industry and research
- Demonstrate the ability to obtain, process and clean data for analysis
- Use existing tools to visualise and analyse data
- Formulate and test theories about data
- Communicate discoveries effectively

What we learn in this course

- **Introduction to data analysis and visualisation**
- **Data pre-processing**
  - Harmonising
  - Missing data
  - Feature & Sample evaluation
- **Projection & reducing dimensions of data**
  - PCA (Principal Component Analysis)
  - NMF (Non-negative Matrix Factorisation)
  - t-SNE (t-Stochastic Neighbour Embedding)
- **Summarising and visualising data**
  - Unsupervised learning (clustering) methods
- **Supervised learning (an introduction)**
  - Support vector machine classifier
- **Interactive graphics with Shiny**
  - A practical web-based application

Textbooks and resources

- **R Programming**
  - An Introduction to R - https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
  - R for Data Science - https://r4ds.had.co.nz/index.html
  - Shiny - https://rstudio.github.io/shiny/tutorial/#welcome
  - GitHub repository - https://github.com/RRafiee/Data-Analysis-and-Visualisation
- **Web references and articles**
  - **PCA**:
    - A step by step explanation of Principal Component Analysis by Zakaria Jaadi; https://builtin.com/data-science/step-step-explanation-principal-component-analysis
    - PCA helps you interpret your data, but it will not always find the important patterns by Jake Lever, Marin Krzywinski & Naomi Altman, Nature Methods 14, pages641–642 (2017); https://www.nature.com/articles/nmeth.4346
  - **NMF**: Learning the parts of objects by non-negative matrix factorization by Daniel D. Lee & H. Sebastian Seung, ***Nature*** volume 401, pages 788–791 (1999); https://www.nature.com/articles/44565

Textbooks and resources

- **t-SNE**: Visualizing Data using t-SNE by Laurens van der Maaten & Geoffrey Hinton; ***Journal of Machine Learning Research***, *Volume* 9, pages 2579-2605 (2008); https://lvdmaaten.github.io/tsne/
- **Pattern Recognition** by Sergios Theodoridis, Konstantinos
  - Clustering
  - Supervised learning methods (SVM & Random Forest)
- **For further study - and not limited to -:**
  - Pattern Recognition and Machine Learning by Christopher M. Bishop
  - Pattern Classification by Richard O. Duda; Peter E. Hart; David G. Stork
  - Introduction to Machine Learning by Ethem Alpaydin

Where (Lectures and Practical)?

- Monday **9:00-10:00** (BEng/MEng CS, BEng/MEng SE, BIT & CIT) at CSB/02/027

Practical

- Tuesday **12:00-14:00** (BEng/MEng CS, BEng/MEng SE) at CSB/01/020
- Thursday **13:00-15:00** (BIT & CIT) at CSB/G/028

- Thursday **15:00-16:00** (BEng/MEng CS, BEng/MEng SE, BIT & CIT) at CSB/02/027
- Friday **11:00-12:00** (BEng/MEng CS, BEng/MEng SE, BIT & CIT) at CSB/02/027

Module content (lecture)

Week	Date	Topic	Notes
1		Introduction to "data" & "data analytics"	
2		Data preparation & pre-processing	Plus R programming
3		Data preparation & pre-processing	Plus R programming
4		Interactive graphics with Shiny	
5		Projection & reducing dimensions of data	
6		Projection & reducing dimensions of data	
7		Summarising and visualising data using clustering	
8		Summarising and visualising data using clustering	
9		Supervised learning	
10		Supervised learning	
11		Interactive graphics with Shiny	
12		Revision	

Module content (practical)

Week	Assignments	Topic	Notes
1	**Practical assignment 1**	Introduction to R programming	Familiarity with R by examples
2			
3	**Practical assignment 2**	Set up a practical medical dataset Data preparation & pre-processing	R programming
4			Addressing missing data
5	**Practical assignment 3**	Projection & reducing dimensions of data	PCA & NMF
6			t-SNE
7	**Practical assignment 4**	Unsupervised learning (Clustering)	K-means & HC
8			AP & GMM
9	**Practical assignment 5**	Supervised learning & classification	SVM
10			Random Forest
11	**Practical assignment 6**	Interactive graphics with Shiny	Web-based application
12			

Practical assignments

Week	Assignments	Mark	Deadline (Date Due)
1	Practical assignment 1		
2		Pass-Fail*	Week 2
3	Practical assignment 2	Pass-Fail	
4			Week 4
5	Practical assignment 3	Pass-Fail	
6			Week 6
7	Practical assignment 4	Pass-Fail	
8			Week 8
9	Practical assignment 5	Pass-Fail	
10			Week 10
11	Practical assignment 6	Pass-Fail	
12			Week 11

*Students must complete the assignment at each lab session to get a "Pass" mark for that lesson

Assessment

Week	Assignments	Mark	Deadline (Date Due)
1			
2			
3	Individual coursework 1		
4		20%	18/10/19 (23:59)
5			
6			
7	Individual coursework 2		
8		40%	15/11/19 (23:59)
9			
10			
11	Individual coursework 3		
12		40%	09/12/19 (23:59)

Discussion and questions

Be Focused, passionate and driven