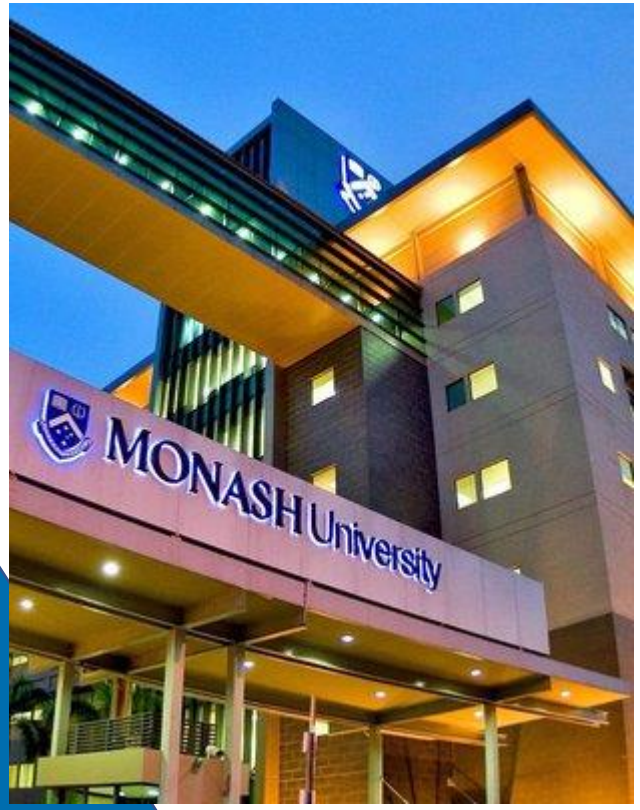
FIT1043 Exam Information



MONASH University

Exam

Content

Lectures and Tutorials

- Everything discussed in the lectures is examinable.
- Python, R, Unix Shell are examinable
- Content of the tutorials explains concepts from the slides

MONASH University

Format of Exam

What will the exam look like?

Exam consists of two parts:

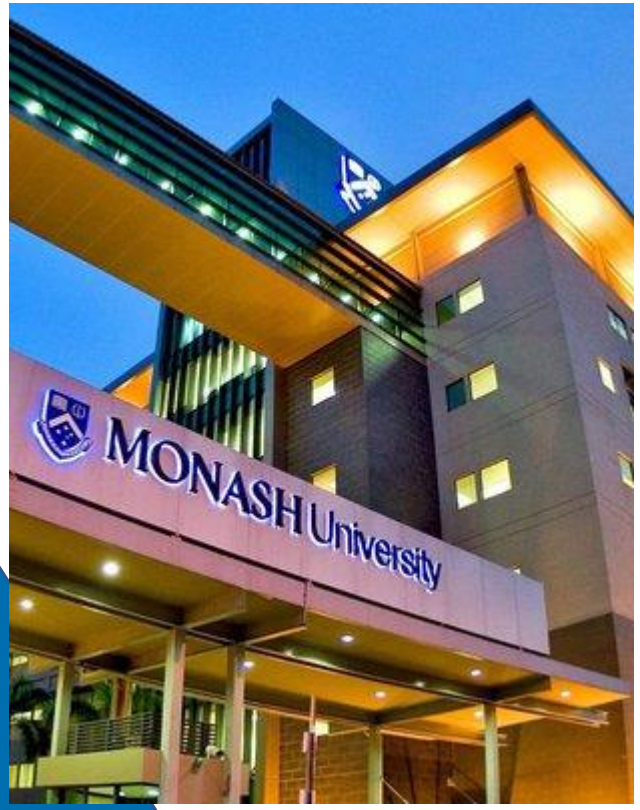
- Section A : Multiple-choice questions (15 questions, 15 marks)
- Section B: Short answer question (25 questions, 2 marks each, total = 50 marks)

Exam duration: 2 hours writing time

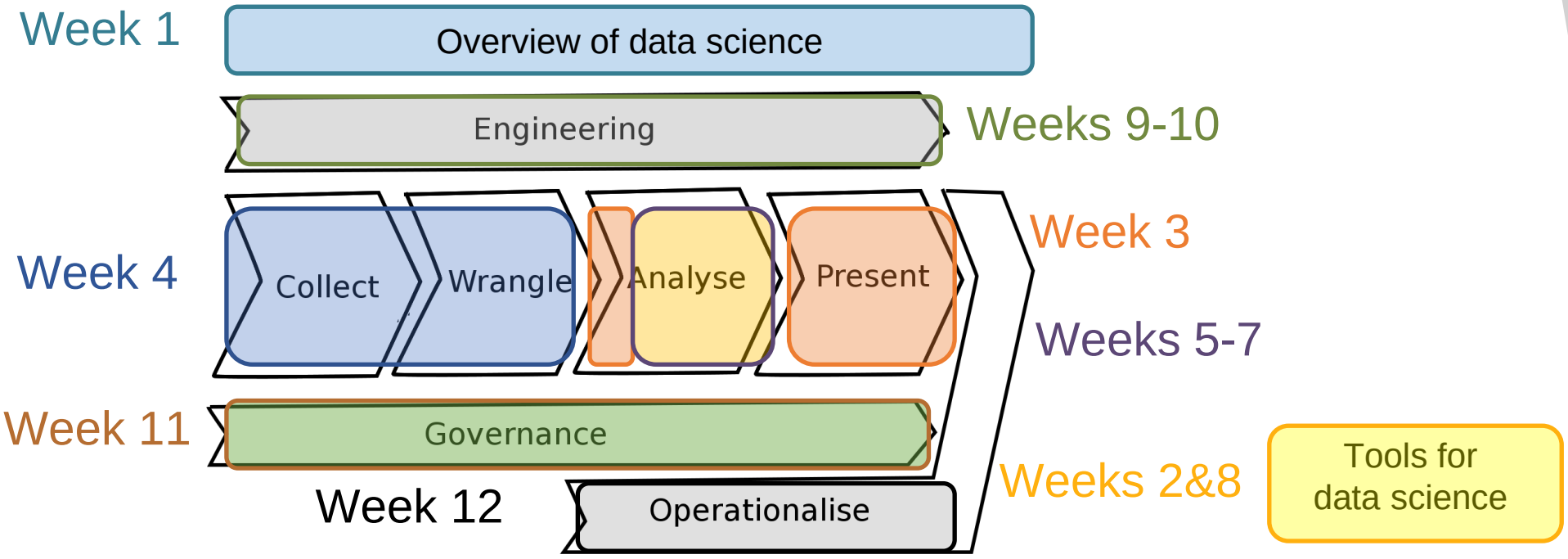
- Reading time: 10 minutes
- Close book

MONASH University

FIT1043 Unit Review



MONASH University



MONASH University

Week 1

Overview of Data Science

What is data science

Drew Conway's Venn diagram

Usefulness of machine learning

Different components of a data science process

Differentiate data science from other related disciplines

https://commons.wikimedia.org/wiki/File:Python-logo-notext.svg

Week 2

Overview of Data Science (con't) and Tools (Python)

Essentials for coding in Python for data science

Interpret given Python codes

Different data science roles and skills and comprehend the differences between them

Impact of data science

MONASH University

Week 3

Visualisation

The importance/power of data visualization

Approaches for data visualisation, explain where each approach is appropriate to be used

Concepts in descriptive statistics

More sophisticated group-by operations in Python

Meme, futurama

Week 4

Collect and Wrangle

Open data and linked open data

How to access to new data sources through APIs

How different APIs work

Data quality problems in datasets

Data wrangling commands in Python

MONASH University

Week 5

Analyse (Models)

What are models and predictive models

Analyse predictive models in different examples

How to evaluate predictive models

How to estimate linear regression model

Linear regression and polynomial regression in Python

"Our data analysis experts can't read your minds. You're going to reach your own decision regarding hiring us, even though it's the same decision we knew in advance you'd make."

https://www.tedgoff.com/

MONASH University

Week 6

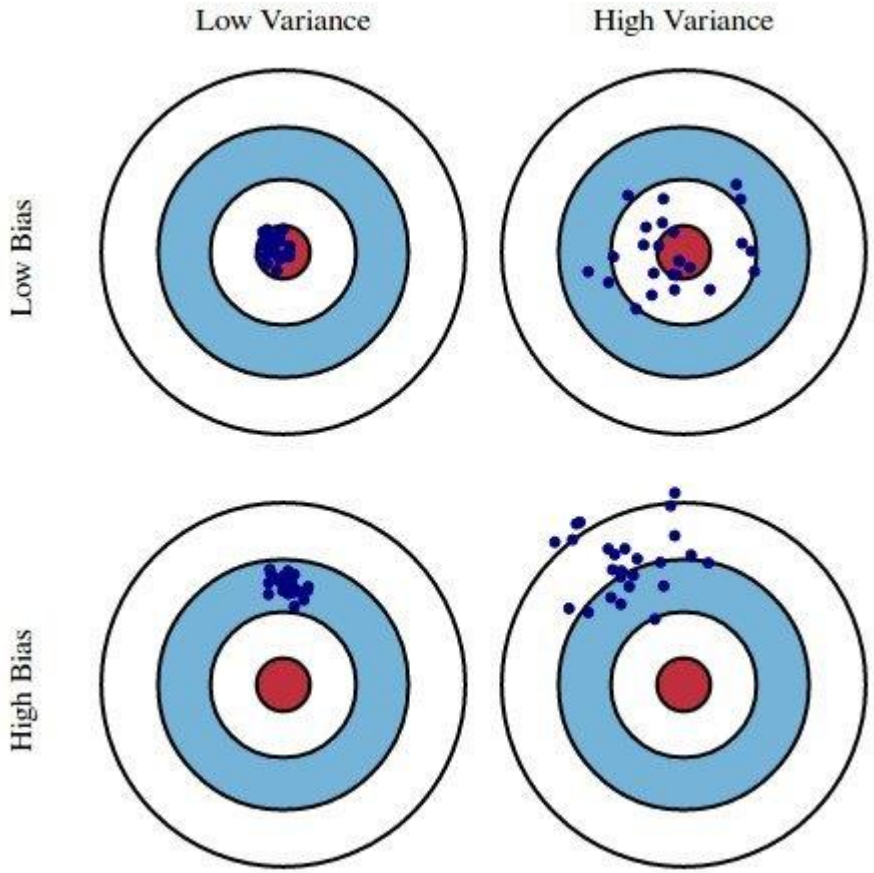
Analyse

Overfitting and underfitting of different models

Bias and variance trade-off

“No Free Lunch Theorem”

What are ensemble models



MONASH University

Week 7

Classification and Clustering

Differentiate between classification and regression models

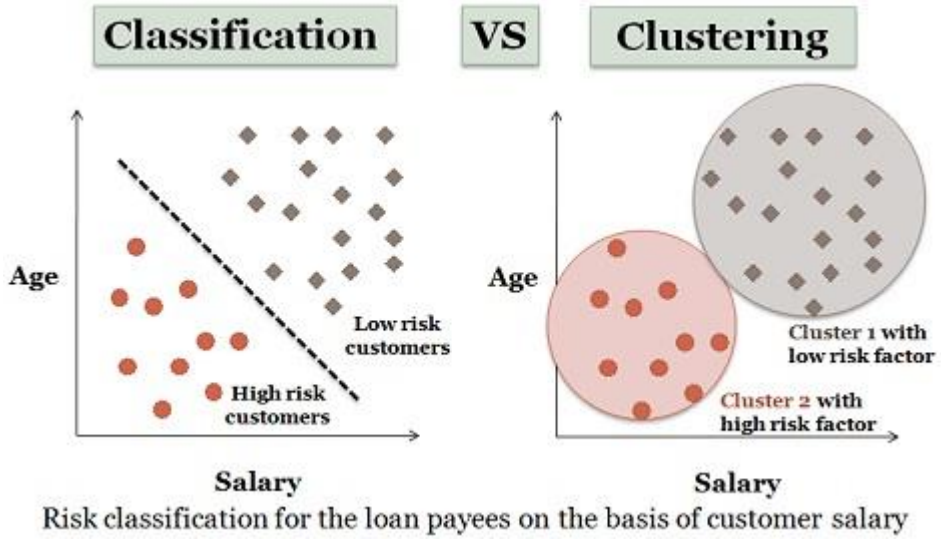
How decision trees and regression trees work

How random forest works

How k-means clustering works

Confusion matrix and prediction accuracy

Different classification metrics



Risk classification for the loan payees on the basis of customer salary

https://techdifferences.com/difference-between-classification-and-clustering.html

MONASH University

https://en.wikipedia.org/wiki/File:R_logo.svg

Week 8

Tools (R)

Essentials for coding in R for data science

Explain and interpret given R commands

Apply R commands for data wrangling, visualisation, exploration and analysis

MONASH University

Week 9

Data Engineering

Characterising big data:

- Volume, Velocity, Variety, Veracity

What is metadata?

- Different types of metadata

Growth laws related to big data:

- Moore’s law, Koomey’s law, Bell’s Law and Zimmerman’s Law

Introduction to Unix Shell commands for data science

MONASH University

Week 10

Data Engineering

Processing big data

- Different types of databases (SQL, graph, noSQL, etc.)
- Different types of processing (interactive, streaming, batch)
- Distributed processing (map-reduce, hadoop, spark, etc.)

What is deep learning and reinforce learning

"Sweetheart, my neural net predicts that you and I are 98.9% compatible. Will you be my Valentine?"

https://www.tedgoff.com/

MONASH University

Week 12

Data Governance

Data management

Confidentiality and privacy

Regulatory compliance

MONASH University

Week	Activities	Assignments
1	Overview of data science	
2	Introduction to Python for data science	
3	Data visualisation and descriptive statistics	
4	Data sources and data wrangling	
5	Data analysis theory	Assignment 1
6	Regression analysis	
7	Classification and clustering	
8	Introduction to R for data science	
9	Characterising data and "big" data	Assignment 2
10	Big data processing	
11	Guest Lecture	
12	Data Governance	Assignment 3

MONASH University

Unit 1043

I hope you've enjoyed the unit and more importantly, learned something.

Do consider follow-on units, where you'll learn the full stuff:

- FIT2086 Modelling for data analysis
- FIT3179 Data visualisation
- FIT3152 Data analytics
- FIT3080 Intelligent systems
- FIT3181 Deep learning
- FIT3182 Big data management and processing

Good luck for your revision and the exam!

MONASH University