Course Overview

ACTL3142 & ACTL5110 Statistical Machine Learning for Risk Applications
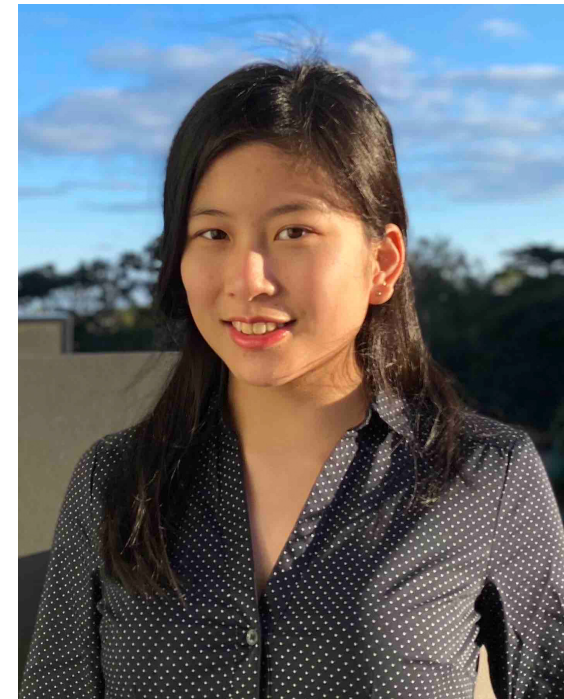
Teaching Team

Patrick Wong LIC for ACTL3142

Patrick Laub LIC for ACTL5110

Jaime Cheung Head Tutor

Tutors: Esme Gao, Tetian Madfouni, Yanfeng Li, Terry Peng, Xi Xin.

UNSW
SYDNEY

Course aims

What will I get out of taking this course?

- Practical skills for building predictive models

- Understanding how they work, when they don't work, interpreting their results

AI's image of you after after taking this course.

Source: DALL-E generated image.

Course Contents

The schedule for each week's topics are:

Regression theory

Lecturer: Patrick Wong

1. Intro. To Statistical Learning
2. Linear Regression I
3. Linear Regression II
4. Logistic Regression
5. Generalised Linear Models

Machine learning

Lecturer: Patrick Laub

6. *Flexibility Week*
7. Machine Learning
8. Moving Beyond Linearity
9. Tree-based Methods
10. Unsupervised Learning

Lecture Outline

- **Learning activities**

- Assessment

- Past Student Feedback

Learning activities

The learning activities of this course involve the following (besides additional self-revision):

1. Self-study:

- Performing reading of relevant textbook chapters

- Doing lab questions (conceptual and applied)

2. Lectures:

- Engaging in each week's lectures

3. Labs:

- Preparing for each week's lab

- Actively engaging in the lab sessions

Course website

At https://unsw-risk-and-actuarial-studies.github.io/ACTL3142/

The ACTL3142 & ACTL5110 course website

Also available by https://laub.au/ml

Course textbook

James, G., Witten, D., Hastie, T., Tibshirani, R., An Introduction to Statistical Learning with Applications in R, Springer, 2nd version, 2021

- Book

  - Electronic copy

  - R labs with detailed explanations

  - A lot of resources including crowd-sourced solutions to questions

- We will cover most of the material in this book.

- Focus on intuition and practical implementation

Springer Texts in Statistics

Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani

An Introduction to Statistical Learning

with Applications in R

Second Edition

Springer
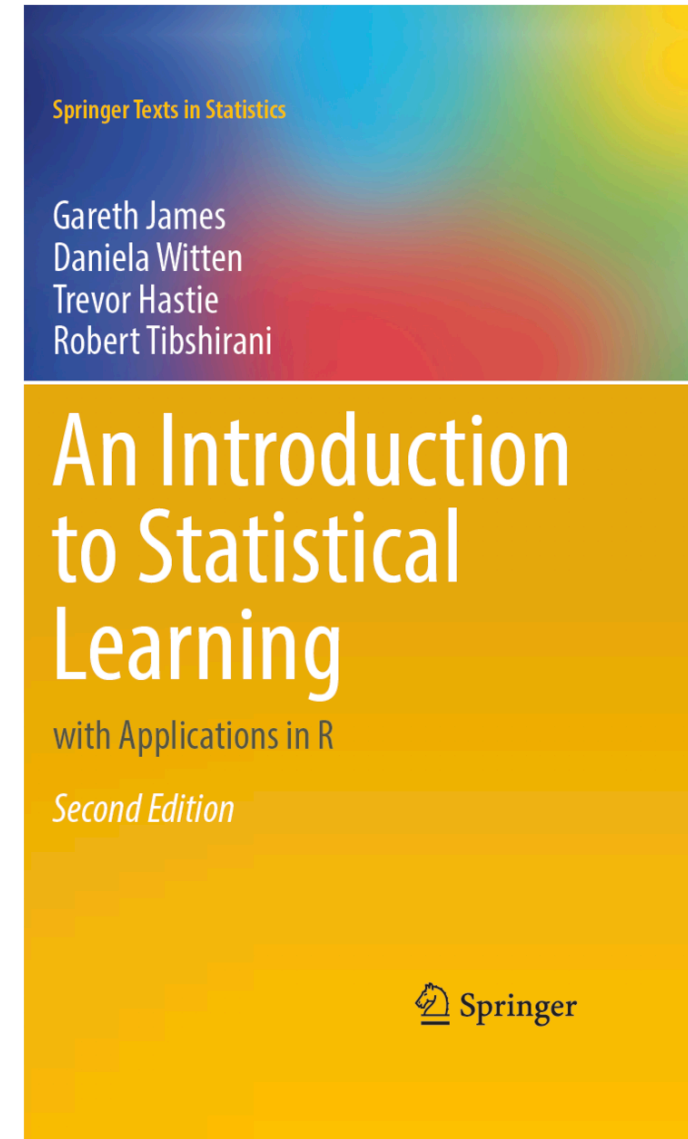
Reading list

Check the course website for the reading list. Currently it starts with:

Week	Required	Optional
1	James et al. (2021): Chapter 1, Chapter 2, Chapter 3 up to and including 3.1.1 Estimating the coefficients	-
2	James et al. (2021): Rest of Chapter 3, Chapter 6.1	-
3	James et al. (2021): Chapters 4.1, 4.2, 4.3, 4.6, 4.7.1, 4.7.2, 4.7.6, 4.7.7	De Jong & Heller (2008): Chapter 4
4	De Jong & Heller (2008): Chapters 5.1, 5.2, 5.3, 5.4	De Jong & Heller (2008): Chapter 5.5
5	De Jong & Heller (2008): Chapters 5.6, 5.7, 5.9, 5.11	De Jong & Heller (2008): Chapter 5.10; Haberman & Renshaw (1996)

Lectures & Labs

Lectures

For ACTL3142 students: The lectures will be 2 hours each week (**not 3**). We will typically not use the third hour on Wednesday **unless exceptional circumstances require us to do so**.

For ACTL3142 students: There won't be a Monday lecture in Week 3 due to the public holiday! A recording will be made available on Moodle instead for this week.

On Wednesdays from 9-10 am we will hold **weekly online consultation** on Zoom (see Moodle for the link).

There won't be a Monday lecture in Week 3 due to the public holiday!

Labs

They're on the course website.

Have a go at them yourself, after the lecture but before your tutorial.

Ed Forum I

The forum will be the course's primary mode of communication. Questions about the theory, the running of the course, or the assignments should be asked on the Ed forum. You have to **open the Ed forum at least once** in order to be enrolled and thus to get any announcements. Don't ask personal questions on the forum. Email those questions to the head tutor who may escalate them to the lecturers.

The Ed forum is accessible via the Moodle page

Ed Forum II

The benefits of asking questions on our Ed forum:

1. It reduces the likelihood of repetitive answers being written. If you have a question, very likely someone else in the course may have the exact same question you would like to ask!

2. It creates a community. At least this way you can see other students will also have questions and that you are not struggling alone.

3. It gives you the opportunity to help others. If you know the answer, please answer! One of the best ways to solidify your understanding is to explain your answer to someone else. The forum gives you that opportunity to do so.

4. It allows you to get help at any point in the course. Please don't be shy in asking any question about the course. We are here to help you do well.

5. It allows you to also clarify your understanding of concepts. Forums give you to pose questions in "I think the answer is …" way, which allows us to give feedback on your thoughts.

Ed Forum III

Before asking a question on the forum, try the following steps:

1. See if your question has already been answered on the forum.

2. Check through the lecture notes / course content to see if you can answer it yourself.

3. Ask in the tutorials / lectures / consultation hours (if they haven't passed already). This allows you to get more immediate feedback as you can ask follow-up questions rather than waiting for us to respond to your follow-up questions.

We only ask you to try first, as we don't want the forum to become your crutch. After all, the forum won't be available in the final exam. If you find someone else's question that you know the answer to, please try to answer it directly. As mentioned before, the forum is designed to be a community and safe space. Please contribute to our community. Plus, you will learn the content very well if you can solve someone else's problem on the Ed forum!

Lecture Outline

- Learning activities

- **Assessment**

- Past Student Feedback

Course Grade Breakdown

1. StoryWall (10%)

2. Project (30%)

3. Exam (60%)

What is StoryWall?

The formative assessments are designed to encourage you to engage with the course content and start experimenting early, without the fear of needing to be 100% correct. You will complete them by submitting posts to specific Moodle forums.

There are two steps for each StoryWall:

1. You must attempt **every component** of the question, and show either your solution or your partial solution to every component, and

2. You must give feedback to another student's submission, and it must be specific feedback. If you write some feedback that could easily be transplanted to any other student's submission (e.g., "Great analysis!" or "Your formatting was very good!") we will treat this component as incomplete.

Read the full instructions on this part of the ACTL3142 Moodle page or on the ACTL5110 Moodle page.

StoryWall Grading

Both components must be completed to get the grade (and there is no half-grades). Take specific note of attempting every component and giving specific feedback. Missing a component or giving non-specific feedback, even if the rest of your submission was very high quality, will result in a 0.

The **comments from your peers will be the primary source of feedback** you will receive on these formative submissions. With that in mind, you should give imagine the kind of feedback you wish to receive from your peers and give that level of quality feedback to someone else.

There are 5 StoryWall tasks, each worth 3% each. The best 4 of 5 being counted, maxing out at 10%. They are due on Friday at 11:55 am in Weeks 3, 5, 7, 9, 10.

You should realistically aim to get 10/10 marks for this…

Project (30%)

Part 1 (9%)

Due Week 5 Friday 11:55 am

This will be a data science task using the techniques you learn from the regression theory half of the course.

Part 2 (21%)

Due Week 9 Friday 11:55 am

This follows on from the previous task using the techniques you learn from the machine learning half of the course.

Due dates

All due dates are at Friday 11:55 am of the following weeks:

1. $\emptyset$

2. $\emptyset$

3. StoryWall 1

4. $\emptyset$

5. StoryWall 2 **and** Project Part 1

6. $\emptyset$

7. StoryWall 3

8. $\emptyset$

9. StoryWall 4 **and** Project Part 2

10. StoryWall 5

Late policy

If submitting late, you must apply for special considerations through UNSW central system. If you ask us for an extension, we will refer you to the special considerations system.

Without special consideration, **late StoryWalls will not be marked**. I have noticed that special considerations will not be granted for StoryWall tasks if you can still get full marks without that task.

For the project, the general policy is:

> Late submission will incur a penalty of 5% per day or part thereof (including weekends) from the due date and time. An assessment will not be accepted after 5 days (120 hours) of the original deadline unless special consideration has been approved.

Example: Late policy for Report Part 2

Report Part 2 (worth 21% course grade) is due Week 9 Friday 11:55 am.

If you submit without special consideration on:

- Week 9 Friday 11:55 am, you have no late penalty.
- Week 9 Friday 11:56 am, you have a **5% penalty**.
- Week 9 Saturday 11:56 am, you have a **10% penalty**.
- Week 10 Sunday 11:56 am, you have a **15% penalty**.
- Week 10 Monday 11:56 am, you have a **20% penalty**.
- Week 10 Tuesday 11:56 am, you have a **25% penalty**.
- Week 10 Wednesday 11:56 am, you will **get 0 marks**.

E.g. a submission on Saturday 11:56 am (10% penalty) which was graded as 80/100, would be recorded as 72/100, and hence an overall course grade of 15.12% out of the maximum 21%.

Special case: Late policy for Report Part 1

However, as a special case **just for Project Report Part 1**, we will not apply the 5% per day penalty for the first 72 hours after the deadline.

Report Part 1 is due Week 5 Friday 11:55 am.

If you submit without special consideration on:

- Week 6 Monday 11:55 am, you have no late penalty.
- Week 6 Monday 11:56 am, you have a **20% penalty**.
- Week 6 Tuesday 11:56 am, you have a **25% penalty**.
- Week 6 Wednesday 11:56 am, you will **get 0 marks**.

Plagiarism and ChatGPT

Plagiarism

Do not send or show your work to another student. You will be penalised along with them!

If you use something directly from an external resource, cite the source.

ChatGPT

You will add a "Generative AI usage" appendix to your reports, detailing how you used AI (what outputs, what prompts).

If you do not use AI, then you will still need the appendix that says that.

Sharing code is definitely cheating.

Source: DALL-E generated image.

Exam

We have many years of past exams for you to practice on, though the last two exams are the most appropriate for the current edition of the course.

In person, invigilated exam, on campus using Inspera.

Source: DALL-E generated image.

Lecture Outline

- Learning activities

- Assessment

- **Past Student Feedback**

Most students enjoyed the course

What were the best things about this course?

"This was one of the best ACTL courses I have ever taken. It has stimulated my interest in data analytics and modelling and I am so glad that the course is structured the way it is. Please do not change anything about the course or the lecturers, they were amazing." (2023 T2)

"the textbook readings were super helpful in getting us to read the concepts in more detail" (2023 T2)

"There were a few things I really loved about this course. The course content wasn't too mathematics focused. It was nice to take a break from all the hardcore calculus. This my first ever machine learning course and it was very cool to see how machine learning works." (2023 T2)

Many enjoyed how practical this course was

"The practical application of techniques on real data was stimulating." (2023 T2)

"The content was really interesting and very practical. Really sparked my interest in data science and modelling in general." (2023 T2)

"Refreshing focus on the theoretical and application aspects and constructions of models rather than mathematical/numerical focus." (2023 T2)

"I really enjoyed practicality of the course, where i felt it is something I could genuinely use in the workforce." (2024 T1)

Some liked the practicality maybe too much

"Reflection: This course has been great! Personally the ACTL3142 course introduced me to statistical machine learning and data analysis techniques used in actuarial work. However, as I delved deeper into the course, I found myself more drawn to the underlying mechanisms of these predictive models and statistical learning techniques. I became more interested in the 'how' rather than the 'what'. This curiosity led me to explore the field of computer science, which forms the backbone of these data analysis techniques.This curiosity led me to explore computer science, which provides a deeper understanding of these tools. This shift in interest from application to creation is why I'm considering changing my degree from actuarial to computer science!" (2023 T2)

Some *loathed* how practical this course was

What could be improved?

"Everything. ACTL courses is meant to be all about applying mathematical theory, we arent compsci people who writes 500+ lines of code per assignment. We are not enrolled in compsci degrees for a reason but now we are forced to do a half theory half compsci course. The theory isnt math heavy either, its more about learning and interpreting models, instead of performing calculations. Even if you want us to do heavy assignments like this one, at least give us R tutorials which cover all the relevant code that we need???? We have compsci's heavy assignment but no spoon feeding code like compsci lectures at all????? Also, 30% for the assignment is way too much, we want the 70% finals instead of 60%." (2023 T2)

Our perspective

This course *will require you to complete a substantial amount of programming*. It is not a computer science course by any measure, but for some students (depending on your background) this will be quite a big step up in the size and difficulty of the coding tasks.

Statistical/machine learning coding is a *very enjoyable component of your whole degree*, though simultaneously it can be one of the *most frustrating parts*. There is no way to progress in your skills without simply *making a lot of mistakes and learning from them*. We recommend that you view this course as a safe environment to learn these skills, and potentially make some mistakes now, so that you have the capability and the confidence to perform make informed decisions later in your career based on rigorous understanding of the data.

Your future career will contain a *sizeable data science / statistical programming element*. This is a key part of your chosen profession, and in fact, if you complete this course and realise that you are truly averse to any coding, then it perhaps worthwhile to reflect on your career path now rather than have expectations which are unsatisfied in the future.

Time management and coding themes

Is there anything you wish you could have told your former self before starting to help them be prepared to learn these topics?

> "As a college returner, I've found the course immensely valuable. It has opened my eyes to the field of machine learning, and I've garnered many ideas while completing the assignments. However, lacking any background in R programming has made adapting to the lab materials quite challenging. If I could offer advice to my past self regarding course preparation, I would suggest learning R first." (2023 T2)

> "If I could have told myself one thing before starting this course, it would have definitely been to manage my time effectively and spend more time on exploring further questions in coding through R as I particularly struggled to manage my workload when working on the assignment and would have benefitted if I started it earlier or had a much better understanding of coding in R." (2023 T2)

Time management and coding themes II

"Something i wish i could have told myself at the start of the term was to truly learn and understand the R codes behind modelling and what each function means." (2023 T2)

"If I could advise my past self on something, I would say brush up on R before beginning the course, and get better at using if functions mainly." (2023 T2)

"If I were to go back in time and tell myself one thing, I would definitely tell myself to not underestimate the time it takes to code these models, as they took significantly longer than I expected - having a deep understanding of how these models work is important, yes, but so is learning how to code them so they run quickly and efficiently in R." (2023 T2)

Time management and coding themes III

"If there was anything I wish I could have told my former self before starting, it would be to start studying and try to stay as up to date as possible as I believed that due to being behind, I missed out on some details that I should have incorporated into the assignment and would cost me marks. Also, I would've told myself to learn how to write a report properly." (2023 T2)

"Furthermore, I would have told myself to keep more on top of the course readings as I had missed some throughout the term. Additionally, I could have spent more time on some content I found difficult such as the unsupervised learning content." (2023 T2)

Get a rubber duck

This is a real thing

Source: Wikipedia, Rubber duck debugging

References

De Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data*. Cambridge University Press.

Haberman, S., & Renshaw, A. E. (1996). Generalized linear models and actuarial science. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 45(4), 407–436.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer.