ETC3250/5250: Introduction to Machine Learning

Getting started

Lecturer: *Professor Di Cook*

Department of Econometrics and Business Statistics

ETC3250.Clayton-x@monash.edu

Week 1

About this class

Meet the team

Me - Hi! Responsible for everything.

Kenji - Tutoring four sessions.

Patrick - Tutoring three sessions.

Shin (data analyst at EY) - Tutoring two sessions.

Where to find things

- Moodle: zoom links, assignments, marks, discussion forum, and announcements
- Website (https://iml.numbat.space) schedule, lecture slides, tutorial instructions
- Textbook (https://www.statlearning.com) free pdf online, data and code (we will use more modern coding style though)
- Got questions? Post in the discussion forum. Of a private nature (e.g. about marks) you can send email to etc3250-clayton-x@monash.edu.

MONASH University

Computing

- Using R (2021-02-15, Lost Library Book) and the RStudio (1.4.1103) interface
- Slides are made in Rmarkdown, xaringan style, using a monash theme made by Emi Tanaka, best viewed as html.
- The code for creating examples in the lecture slides can be found in the Rmd file. Look for the code blocks in the Rmd like

```
library(tidyverse)
library(gapminder)
library(gridExtra)
```

MONASH University

Learning objectives

- Select and develop appropriate models for regression, classification or clustering
- Estimate and simulate from a variety of statistical models, and measure the uncertainty of a prediction using resampling methods
- Manage large data sets in a modern software environment, and explain and interpret the analyses undertaken clearly and effectively
- Apply business analytic tools to produce innovative solutions in finance, marketing, economics and related areas

How to do well in this class

- Come to class
- Work through tutorial exercises
- Bring questions to consultation
- Read ahead in the textbook, and attempt some of the questions
- Start assignments early!
- Ask and answer questions in the discussion forum
- Do the weekly quizzes
- Don't procrastinate - if you don't understand something or can't work how to do a problem, ask, get help, no question is too simple or stupid!

Where this material fits in your career skills development

> *See the **Math and Statistics** box: machine learning, statistical modeling, supervised and unsupervised learning.*

> *And also the **Communication and Visualisation** box: Translate data-driven insights into decisions and actions, plotting using ggplot2.*

If you like this course, what others could you do

ETC3555/5555 - Statistical Machine Learning: This unit covers the methods and practice of statistical machine learning for modern data analysis problems. Topics covered will include recommender systems, social networks, text mining, matrix decomposition and completion, and sparse multivariate methods. All computing will be conducted using the R programming language.

ETC3580/5580 - Advanced Statistical Modelling: This unit introduces extensions of linear regression models for handling a wide variety of data analysis problems. Three extensions will be considered: generalised linear models for handling counts and binary data; mixed-effect models for handling data with a grouped or hierarchical structure; and non-parametric regression for handling non-linear relationships. All computing will be conducted using R.

Introduction to machine learning

Notation 1/6

n number of observations or sample points

p number of variables or the dimension of the data

A data matrix is denoted as:

$$\mathbf{X}_{n\times p} = (\mathbf{x}_1\ \mathbf{x}_2\ \dots\ \mathbf{x}_p) = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

Notation 2/6

The i^{th} observation is denoted as

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

The j^{th} variable is denoted as

$$x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

Notation 3/6

A transposed data matrix is denoted as

$$\mathbf{X}_{p\times n}^T = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix}$$

and

$$x_i^T = \begin{pmatrix} x_{i1} & x_{i2} & \dots & x_{ip} \end{pmatrix}$$

Notation 4/6

The response vector, when it exists, is denoted as

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

and then an observation can be written as

$$\mathcal{D} = \{(y_i, x_i)\}_{i=1}^n = \{(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)\}$$

where x_i is a vector with p elements.

Notation 5/6

If $\mathbf{y}$ is categorical, with K levels, it can be useful to write it as a binary matrix

$$\mathbf{Y}_{n\times K} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Matrix multiplication

Suppose that

$$\mathbf{A}_{2\times 3} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

$$\mathbf{B}_{3\times 4} = \begin{pmatrix} -1 & -2 & -3 & -4 \\ -5 & -6 & -7 & -8 \\ -9 & -10 & -11 & -12 \end{pmatrix}$$

then

$$\mathbf{AB}_{2\times 4} = \begin{pmatrix} -38 & -44 & -50 & -56 \\ -83 & -98 & -113 & -128 \end{pmatrix}$$

Pour the rows into the columns.

Note: You can't do $\mathbf{BA}$

Using R as a matrix calculator

```
a <- matrix(c(1,2,3,4,5,6),
            ncol=3, byrow=T)
b <- -1*matrix(c(1,2,3,4,5,6,
                 7,8,9,10,11,12),
               ncol=4,byrow=T)
```

```
a%*%b
```

Inverting a matrix

Suppose that $\mathbf{A}$ is square and symmetric

$$\mathbf{A}_{2\times 2} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$$

then the inverse is (if $ad - bc \neq 0$)

$$\mathbf{A}_{2\times 2}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

and where $\mathbf{A}\mathbf{A}^{-1} = I$

$$\mathbf{I}_{2\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Notation 6/6

$d (\leq p)$ is used to denote the number of variables in a lower dimensional space, usually by taking a projection.

A is a $p \times d$ orthonormal basis, $A^T A = I_d$ ($A'A = I_d$).

The projection of $\mathbf{x}_i$ onto A is $A^T \mathbf{x}_i$

Different types of learning

1. Supervised learning: y_i is available for all x_i
   - Regression: quantitative y_i
   - Classification: categorical y_i

2. Unsupervised learning: y_i unavailable for all x_i

3. Semi-supervised learning: y_i available for some x_i (not covered in this unit)

> Being able to recognise the type of problem is an important skill.

What type of problem is this? (1/3)

Food servers' tips in restaurants may be influenced by many factors, including the nature of the restaurant, size of the party, and table locations in the restaurant. Restaurant managers need to know which factors matter when they assign tables to food servers. For the sake of staff morale, they usually want to avoid either the substance or the appearance of unfair treatment of the servers, for whom tips (at least in restaurants in the United States) are a major component of pay.

In one restaurant, a food server recorded the following data on all customers they served during an interval of two and a half months in early 1990. The restaurant, located in a suburban shopping mall, was part of a national chain and served a varied menu. In observance of local law the restaurant offered seating in a non-smoking section to patrons who requested it. Each record includes a day and time, and taken together, they show the server's work schedule.

What type of problem is this? (2/3)

Measurements on rock crabs of the genus *Leptograpsus*. One species *L. variegatus* had been split into two new species, previously grouped by color, orange and blue. Preserved specimens lose their color, so it was hoped that morphological differences would enable museum specimens to be classified. There are 50 specimens of each sex of each species, collected on site at Fremantle, Western Australia. For each specimen, five measurements were made, using vernier calipers.

What type of problem is this? (3/3)

This data contains observations taken from a high-energy particle physics scattering experiment that yielded four particles. The reaction $\pi_b^+ p_t \to p \pi_1^+ \pi_2^+ \pi^-$ can be described completely by seven independent measurements. Below, $\mu^2(A,$ and $\mu^2(A, B,$ where E and P represent the particle's energy and momentum, respectively, as measured in billions of electron volts. The notation $(p)^2$ represents the inner product $P'P$. The ordinal assignment of the two π^+'s was done randomly. What are the clusters in the data?

Lecturer: *Professor Di Cook*

Department of Econometrics and Business Statistics

ETC3250.Clayton-x@monash.edu

Week 1