Monday May 19

Session 1. Get Data

1.1 Learning Outcomes

A successful student shall be able to:

- *Explain different sources of data.*
- *Explain different types of data formats.*
- *Load data from open data sources into R.*

1.2 Lecture/Demonstration

- *Introduction of the Data Science Life Cycle.*
- *Introduction to RStudio in Posit Connect*
- *Sources of data to include Open data.*
- *Different types of data formats: .csv files, excel files, urls, compressed data, Arrow-Parquet*
- *Rectangular data, vectors and data frames in R*
- *Methods for getting rectangular data from other sources or files into R data frames.*
- *Viewing data and getting summary statistics about data in R data frames.*

1.3 Exercises

- *Exercise on loading data from various sources into R and getting summary statistics.*

Session 2. Clean and Reshape Data

2.1 Learning Outcomes

A successful student shall be able to:

- *Explain the need to clean data.*
- *Clean and shape rectangular data: filter rows, select columns, rename columns, …*
- *Explain basic methods for reshaping data: pivots and joins.*

2.2 Lecture/Demonstration

- *Cleaning data*
- *Filtering rows using logical comparisons and* `%in%`*.*
- *Selecting columns using names and tidyselect.*
- *Reshaping data using pivots.*

2.3 Exercises

- *Exercise on cleaning data in R - filtering, selecting. renaming*
- *Example on reshaping data*

==========

Tuesday May 20

Session 3 Visualize Univariate Data

3.1 Learning Outcomes

A successful student shall be able to:

- *Create histograms, density plots, and bar plots.*
- *Customize plot titles, axis labels, …*

3.2 Lecture/Demonstration

- *Variable types: continuous vs categorical*
- *Plots for continuous variables*
- *Plots for categorical variables*
- *Customizing plots*
- *Example of dynamic titles*

3.3 Exercises

- *Exercise on making and customizing univariate plots.*

Session 4. Visualize Multi-variate Data

4.1 Learning Outcomes

A successful student shall be able to:

- *Create bivariate point plots and box plots.*
- *Use plot aesthetics to code by additional variables.*
- *Add and interpret linear and non-linear smoothers.*

4.2 Lecture/Demonstration

- *Create bivariate point plots and box plots.*
- *Use plot aesthetics to code by additional variables.*
- *Add and interpret linear and non-linear smoothers.*

4.3 Exercises

- *Exercise on plotting bivariate and multi-variate data in R.*

Wednesday May 21

Session 5. Statistical Tests and Models

5.1 Learning Outcomes

A successful student shall be able to:

- *Explain the purpose for statistical tests and a Null hypothesis.*
- *Use and interpret the t.test function, aov/anova and lm functions in R.*
- *Generate residual plots in R.*

5.2 Lecture/Demonstration

- *Why statistical tests and understanding a Null hypothesis.*
- *Interpreting a p-value.*
- *Using and interpreting the t.test function, aov/anova, and lm functions in R.*
- *Generating and interpreting residual plots in R.*

5.3 Exercises

- *Exercises on t.test, linear regression from a scatter plot, and analysis of variance from Box Plot.*

Session 6. Classification Models

6.1 Learning Outcomes

A successful student shall be able to:

- *Explain the difference between regression and binary classification.*
- *Use the glm function to generate and interpret a logistic regression.*
- *Explain False Positives and False Negatives from a confusion matrix.*

6.2 Lecture/Demonstration

- *Binary Classification*
- *Logistic Regression*
- *Confusion Matrices and metrics*

6.3 Exercises

- *Exercise on generating a logistic regression model and interpreting results.*

Thursday May 22

Session 7. Overfitting and Bias-Variance Tradeoff

7.1 Learning Outcomes

A successful student shall be able to:

- *Explain the danger of overfitting.*
- *Explain the bias -variance trade off.*
- *Explain the purpose of training and test data sets.*

7.2 Lecture/Demonstration

- *Degrees of Freedom and Restrictive versus flexible models*
- *Over-fitting*
- *Bias-Variance Trade off*

- *Validating and models by splitting data.*

7.3 Exercises

- *Exercise on using splines to explore degrees of freedom and restrictive vs flexible models.*
- *Exercise on using training and test data for linear regression.*

Session 8. Evaluating and Tuning Models

8.1 Learning Outcomes

A successful student shall be able to:

- *Explain the need to evaluate and tune models.*
- *Consider strategies for variable/feature selection - Regularization.*
- *Explain strategies for evaluating model performance.*

8.2 Lecture/Demonstration

- *Optimizing Loss Functions*
- *Evaluation Metrics*
- *Variable Selection and Partial F-Tests*
- *ROC Curves in Classification*

8.3 Exercises

- *Exercise on variable selection and tuning in regression and classification.*

Friday May 23

Session 9. Neural Networks

9.1 Learning Outcomes

A successful student shall be able to:

- *Explain the basic architecture of a neural network.*

- *Explain the role of back-propagation and gradient descent.*
- *Build a basic neural network using R.*

9.2 Lecture/Demonstration

- *Convergence of big data (cheap storage and memory), big processing power (GPUs), and algorithmic advances.*
- *Neural Network architecture*
- *Role of Activation Functions and gradient descent and back-propagation.*
- *Challenges in Over fitting and convergence*

9.3 Exercises

- *Exercise on building a neural network (neural net library)*

Session 10. Generative AI and Prompt Engineering

10.1 Learning Outcomes

A successful student shall be able to:

- *Explain what makes generative AI different.*
- *Explain the concepts of tokens and embeddings.*
- *Explain the concept of semantic similarity and prediction.*
- *Apply concepts in prompt engineering.*

10.2 Lecture/Demonstration

- *Generative AI compared to predictive models.*
- *Tokenization and Embedding*
- *Semantic similarity for prediction.*
- *Dangers of overfitting and the need for regularization.*
- *Elements of prompt engineering.*

10.3 Exercises