Week 2 Pass Task – SIT720 – Damon Vizl – s223545885

Summary

* We covered some math concepts related to machine learning including:
  + Statistics and Probability. We discussed how to calculate basic probability, joint probability, conditional probability and Bayes rule.
  + Normal distribution, uniform distribution and Bernoulli distribution.
  + We also touched on the central limit theorem that populations tend to a normal distribution as the sample size grows.
* We discussed data wrangling methods as well including an understanding of features, matrices, vectors, signals and data.
* We discussed encoding (i.e. transforming categorical data into numbers to allow the ML model to process it)
* We discussed scaling and normalization to ensure that the size of a certain feature doesn’t oversaturate the model and skew the model.

Reading list and Knowledge Reflection

In order to enhance my understanding of this weeks subjects I read more information about the data wrangling methods using python extensions numpy, pandas, sklearn and scipy. I went to the sklearn website and perused the API to further understand what the encoding was doing, specifically the One Hot Encoder that I used in the problem set.

I also delved deeper into the scipy stats module to learn about the Shaprio-Wilks test to confirm if a data set is normally distributed.

Quiz Results

Graphical user interface, text, application, email

Description automatically generated