Introduction to Statistical Learning

Signature ↓

Initials and Surname		
Student number		

Duration: $3\frac{1}{2}$ hours Full marks: **[50]**

All answers should be properly discussed and interpreted.

Question 1: Monte Carlo integration [13]

Consider the following population growth model:

$$\begin{aligned} P_t &= \alpha(1 - \beta\, e^{-\gamma t}) \\ &= 5160(1 - 0.8\, e^{-0.1t}) \end{aligned}$$

with P_t the population size in period t, $\alpha > 0$, $0 < \beta < 1$ and $0 < \gamma < 1$.

1. Calculate $\int_2^{20} P(t)dt$ using Monte Carlo integration. (7)
2. What is the maximum value of the function, P_t? (3)
3. Interpret the parameters α, β and γ in the model above. (3)

Question 2: $k - nn$ modelling [37]

Consider the attached data, *q2.xlsx*.

1. How many reponse levels (variable Y) are there? (2)
2. Fit a k nearest neighbour classification model. Adjust the classification rule to be a distance weighted classification rule. That is the majority vote classification rule is amended to be a weighted average based on the distance of observations from the point of interest. Use the following weight function:
$$w_{ji} = \frac{1}{d(obs_j, obs_i)}$$
with $j = 1, 2, \ldots, n$, $i = 1, 2, \ldots, k$ and $d(obs_j, obs_i)$ the distance between observation j and the i^{th} nearest neighbour. Do not use cross validation or a test data set (15)
3. Use k-fold (10-fold) cross validation to determine an appropriate value for k (the number of neighbours) used in the modelling. (10)

   The k in k-fold cross validation is not the same as the k in $k - nn$ modelling
4. Determine a 95% confidence interval for the value of $count - R^2$, for your chosen value of k. HINT: $count - R^2$ is the proportion of observations that are classified correctly. (10)

Examination instructions

1. This is a not a group assignment.
2. You may make use of any appropriate material to help you complete the paper. You may not share this question paper or discuss it with anyone.
3. You have $3\frac{1}{2}$ hours to complete the test paper.
4. Answer all the questions clearly in **one** PDF file. Attach your SAS code as an Appendix to your PDF file.
5. Clearly indicate question numbers next to your answers.
6. Your hand-in will be checked for plagiarism with dedicated software.
7. Submit your PDF via the Turnitin on ClickUp.
8. Submit your PDF document, including your SAS programmes in an appendix. No additional files of any sort should be submitted. Do not submit files in any other format other than PDF. Failure to follow any of these instructions will result in a **zero** mark for the test. Upload the PDF file, named s99999999.pdf, where 99999999 is your student number. **Ensure that your name and student number are clearly indicated on the front page of your hand-in.**
9. There will be 30 minute ClickUp test at the end of the test (included in the $3\frac{1}{2}$ hours allocation).