



**Thursday 13 May 2021**  
**Available from 14:00 BST**  
**Expected Duration: 2 hours**  
**Time Allowed: 4 hours**  
**Timed exam within 24 hours**

**DEGREES OF MSci, MEng, BEng, BSc, MA and MA (Social Sciences)**

# **COMPUTATIONAL SOCIAL INTELLIGENCE H**

## **COMPSCI 4080**

**(Answer all 3 questions.)**

**This examination paper is an open book, online  
assessment and is worth a total of 75 marks**

1. The following table shows the result of a study about the amount of time spent on social media every week:

	Age < 30	Age $\geq$ 30
T > 12 hours	125	32
T $\leq$ 12 hours	79	77

The study participants have been split into two categories:

- People that are less than 30 years old (column “Age < 30”)
- People that are at least 30 years old (column “Age  $\geq$  30”).

These two categories have been further been split into two subgroups:

- People that spend more than 12 hours per week on social media (row “T > 12 hours”)
- People that spend at most 12 hours per week on social media (row “T  $\leq$  12 hours”).

- (a) Use the data above to build a contingency table.

[5 marks]

- (b) Write the formula of a  $\chi^2$  variable that can test the following research hypothesis: “People younger than 30 and people at least 30 years old tend to spend a different amount of time per week on social media”.

[5 marks]

- (c) Provide the null hypothesis corresponding to the research hypothesis stated at point (b) of this question.

[5 marks]

- (d) Write the formula of the expectation values for the relevant cells of the contingency table.

[5 marks]

- (e) Calculate the number of degrees of freedom for the  $\chi^2$  variable you have defined at point (b) of this question (write the formula you use it to calculate it).

[3 marks]

- (f) Calculate the value of the  $\chi^2$  variable defined at point (b) of this question. Compare the value of the  $\chi^2$  variable with the critical value 3.841 (corresponding to confidence level  $\alpha = 0.95$ ).

[2 marks]

- (g) Explain whether the null hypothesis can be rejected and whether this means that the research hypothesis is true.

[5 marks]

2. Consider a classification problem where there are two classes  $C_1$  and  $C_2$  with a-priori probabilities  $p(C_1) = 0.67$  and  $p(C_2) = 0.33$ , respectively. Consider 5 different feature vectors ( $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$  and  $\bar{x}_5$ ) such that the likelihoods are as follows:

$p(\bar{x}_1 C_1) = 0.50$	$p(\bar{x}_1 C_2) = 0.10$
$p(\bar{x}_2 C_1) = 0.30$	$p(\bar{x}_2 C_2) = 0.20$
$p(\bar{x}_3 C_1) = 0.10$	$p(\bar{x}_3 C_2) = 0.30$
$p(\bar{x}_4 C_1) = 0.07$	$p(\bar{x}_4 C_2) = 0.35$
$p(\bar{x}_5 C_1) = 0.03$	$p(\bar{x}_5 C_2) = 0.05$

- (a) Explain how the prior rule works.

[4 marks]

- (b) Find the class of the five vectors  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$  and  $\bar{x}_5$  according to the prior rule.

[6 marks]

- (c) Explain how the posterior rule works.

[8 marks]

- (d) Find the class of the five vectors  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$  and  $\bar{x}_5$  according to the posterior rule.

[12 marks]

**3.** Explain the most important aspects of one of the four experimental studies that have been presented during the course:

1. Synthetic impressions;
2. Speech and Personality;
3. Facial Expressions;
4. Nonverbal Behaviour in Phone Calls.

The answer must include no more than 1,000 words. The text beyond the first 1,000 words will not be considered.

[15 marks]