R249/419C

DUBLIN INSTITUTE OF TECHNOLOGY
KEVIN STREET, DUBLIN 8

BSc. (Honours) Degree in Information Systems / Information Technology

Stage 4

SUPPLEMENTAL EXAMINATIONS 2015

ARTIFICIAL INTELLIGENCE II [CMPU4011]

Dr. John Kelleher
Dr. Deirdre. Lillis
Mr. P. Collins

Duration: 2 Hours

Question 1 is **compulsory**

Answer Question 1 (40 marks) **and**

any 2 Other Questions (30 marks each).

1. (a) Explain what is meant by **inductive learning**.

(5 marks)

(b) In the context of machine learning, explain what is meant by the term **inductive bias** and illustrate your explanation using examples of inductive biases used by machine learning algorithms.

(15 marks)

(c) Table 1 shows the predictions made for a categorical target feature by a model for a test dataset.

(i) Create the **confusion matrix** for the results listed in Table 1.

(5 marks)

(ii) Calculate the **classification accuracy** for the results listed in Table 1.

$$classification\ accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

(5 marks)

(iii) Calculate the **average class accuracy (harmonic mean)** for the results listed in Table 1. (During this calculation you should round all long floats to 4 places of decimal.)

$$average\ class\ accuracy_{HM} = \frac{1}{\dfrac{1}{|levels(t)|}\displaystyle\sum_{l \in levels(t)} \dfrac{1}{recall_l}}$$

(10 marks)

Table 1: The predictions made by a model for a categorical target on a test set of 20 instances

ID	Target	Prediction	ID	Target	Prediction
1	false	false	11	false	false
2	false	false	12	false	true
3	false	false	13	false	false
4	false	false	14	false	false
5	false	true	15	false	false
6	false	false	16	false	false
7	false	false	17	true	false
8	false	false	18	true	false
9	false	false	19	true	false
10	false	false	20	true	true

2. (a) You are building a recommender system for an large online shop that has a stock of over 100,000 items. In this domain the behaviour of individuals is captured in terms of what items they have bought or not bought.

   (i) Table 2 (below) lists 3 different models of similarity that work on binary data, similar to the data in this domain (**Russell-Rao**, **Sokal-Michener**, and **Jaccard**). Given that there are over 100,000 items available in the store which of these models of similarity (**Russell-Rao**, **Sokal-Michener**, or **Jaccard**) is most appropriate for this domain. Give an explanation for your choice.

   (5 marks)

   (ii) Table 4 (on the next page) lists the behaviour of two individuals in this domain for a subset of the items that at least one of the individuals has bought; and, Table 5 (also, on the next page) lists the behaviour of a customer **Q** that you want to generate recommendations for. Assuming that the recommender system uses the similarity metric you selected in Part (i) and that the system will recommend to person Q the items that the person most similar to person Q has already bought but that person Q has not bought, **which item or items will the system recommend to person Q?** Support you answer by showing your calculations and explaining your analysis of the results.

   (10 marks)

Table 2: Similarity Metrics for Binary Data.

$$\begin{aligned}
\text{Russell-Rao(X,Y)} &= \frac{CP(X,Y)}{P} \\
\text{Sokal-Michener(X,Y)} &= \frac{CP(X,Y)+CA(X,Y)}{P} \\
\text{Jaccard(X,Y)} &= \frac{CP(X,Y)}{CP(X,Y)+PA(X,Y)+AP(X,Y)}
\end{aligned}$$

(b) Table 6 (on the next page) lists a data set with of 6 examples described in terms of 3 binary descriptive features (**A**, **B**, and **C**) and a target feature (**Target**). You are asked to create a decision tree model using this data. **Which of the descriptive features will the ID3 decision tree induction algorithm choose as the feature for the root node of the decision tree?** Support you anwer with appropriate calculations and dicussions of your results. Note that Table 3 (below) lists some equations that you may find useful for this question.

(15 marks)

Table 3: Equations from information theory.

$$\begin{aligned}
H(\mathbf{f},\mathcal{D}) &= -\sum_{l \in levels(f)} P(f=l) \times log_2(P(f=l)) \\
rem(\mathbf{f},\mathcal{D}) &= \sum_{l \in levels(f)} \frac{|\mathcal{D}_{f=l}|}{|\mathcal{D}|} \times H(t,\mathcal{D}) \\
IG(\mathbf{d},\mathcal{D}) &= H(\mathbf{t},\mathcal{D}) - rem(\mathbf{d},\mathcal{D})
\end{aligned}$$

Table 4: A dataset showing the behaviour of two individuals in an online shop. A 1 indicates that the person bought the item a 0 indicates that they did not.

Person ID	Item 107	Item 498	Item 7256	Item 28063	Item 75328
A	1	1	1	0	0
B	1	0	0	1	1

Table 5: A query instance from the same domain as the examples listed in Table 4. A 1 indicates that the person bought the item a 0 indicates that they did not.

Person ID	Item 107	Item 498	Item 7256	Item 28063	Item 75328
Q	1	0	1	0	0

Table 6: Dataset for the ID3 Algorithm Question

ID	**A**	**B**	**C**	**Target**
1	1	0	1	C1
2	1	1	1	C2
3	1	0	1	C1
4	0	1	1	C2
5	0	1	0	C1
6	0	1	1	C2

3. Table 7 lists a dataset of books and whether or not they were purchased by an individual (i.e., the feature PURCHASED is the target feature in this domain).

   (a) Calculate the probabilities (to four places of decimal) that a **naive Bayes** classifier would use to represent this domain.

   (18 marks)

   (b) Assuming conditional independence between features given the target feature value, calculate the **probability** of each outcome (PURCHASED=Yes, and PURCHASED=No) for the following book (marks will be deducted if workings are not shown, round your results to four places of decimal)

   2ND HAND=False, GENRE=Literature, COST=Expensive

   (10 marks)

   (c) What prediction would a **naive Bayes** classifier return for the above book?

   (2 marks)

Table 7: A dataset describing the a set of books and whether or not they were purchased by an individual.

ID	2ND HAND	GENRE	COST	PURCHASED
1	False	Romance	Expensive	Yes
3	True	Romance	Cheap	Yes
4	False	Science	Cheap	Yes
10	True	Literature	Reasonable	Yes
2	False	Science	Cheap	No
5	False	Science	Expensive	No
6	True	Romance	Reasonable	No
7	True	Literature	Cheap	No
8	False	Romance	Reasonable	No
9	True	Science	Cheap	No

4. (a) A multivariate logistic regression model has been built to predict the propensity of shoppers to perform a repeat purchase of a free gift that they are given. The descriptive features used by the model are the age of the customer, the average amount of money the customer spends on each visit to the shop, and the average number of visits the customer makes to the shop per week. This model is being used by the marketing department to determine who should be given the free gift. The trained model is

$$\begin{aligned}\text{REPEAT PURCHASE} = &- 3.82398 - 0.02990 \times \text{AGE} \\ &+ 0.74572 \times \text{SHOP FREQUENCY} \\ &+ 0.02999 \times \text{SHOP VALUE}\end{aligned}$$

And, the logistic function is defined as:

$$logistic(x) = \frac{1}{1 + e^{-x}}$$

Assuming that the *yes* level is the positive level and the classification threshold is 0.5, use this model to make predictions for each of the query instances shown in Table 8, on the next page.

(12 marks)

(b) The effects that can occur when different drugs are taken together can be difficult for doctors to predict. A machine learning has been trained to distinguish between dosages of two drugs that cause a dangerous interaction and those that cause a safe interaction. There are just two continuous features in this dataset, DOSE1 and DOSE2, and two target levels, *dangerous* and *safe*. There is a non-linear decision boundary between dangerous and safe interactions and, consequently, the following set of basis functions were defined:

$$\begin{aligned}
&\phi_0(\langle \text{DOSE1}, \text{DOSE2} \rangle) = 1 && \phi_1(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1} \\
&\phi_2(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE2} && \phi_3(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1}^2 \\
&\phi_4(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE2}^2 && \phi_5(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1}^3 \\
&\phi_6(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE2}^3 && \phi_7(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1} \times \text{DOSE2}
\end{aligned}$$

Training a logistic regression model using this set of basis functions leads to the following model:

$$\begin{aligned}
&P(\text{TYPE} = \textit{dangerous}) = \\
&Logistic\big(- 0.848 \times \phi_0(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 1.545 \times \phi_1(\langle \text{DOSE1}, \text{DOSE2} \rangle) \\
&\qquad - 1.942 \times \phi_2(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 1.973 \times \phi_3(\langle \text{DOSE1}, \text{DOSE2} \rangle) \\
&\qquad + 2.495 \times \phi_4(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 0.104 \times \phi_5(\langle \text{DOSE1}, \text{DOSE2} \rangle) \\
&\qquad + 0.095 \times \phi_6(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 3.009 \times \phi_7(\langle \text{DOSE1}, \text{DOSE2} \rangle)\big)
\end{aligned}$$

Use this model to make predictions for the query instances in Table 9 and using these prediction explain whether or not the dosage combinations are likely to lead to a dangerous of safe interaction.

(18 marks)

Table 8: The queries for the multivariate logistic regression question

ID	AGE	SHOP FREQUENCY	SHOP VALUE
A	37	0.72	170.65
B	32	1.08	165.39

Table 9: The query instances for the dosage prediction problem

ID	DOSE1	DOSE2
1	0.50	0.75
2	0.10	0.75
3	-0.47	0.18