S249/S249P/419C

DUBLIN INSTITUTE OF TECHNOLOGY
KEVIN STREET, DUBLIN 8

BSc. (Honours) Degree in Information Systems / Information Technology (Part-time)

Stage 4

SUMMER EXAMINATIONS 2016

ARTIFICIAL INTELLIGENCE II [CMPU4011]

Dr. John Kelleher
Dr. Deirdre. Lillis
Dr. Rem Collier

Monday 9th May 2016
4:00 pm to 6:00 pm

Question 1 is **compulsory**

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

any 2 Other Questions (30 marks each).

1. (a) What is **supervised machine learning**?

(5 marks)

(b) Explain what can go wrong when a machine learning classifier uses the wrong **inductive bias**.

(5 marks)

(c) Table 1, on the next page, shows the predictions made for a categorical target feature by a model for a test dataset. Based on this test set, calculate the evaluation measures listed below.

(i) A **confusion matrix**

(6 marks)

(ii) The **misclassification rate**

(4 marks)

$$misclassification\ rate = \frac{(FP + FN)}{(TP + TN + FP + FN)}$$

(iii) The **precision**, **recall**, and **F$_1$ measure**

(12 marks)

$$precision = \frac{TP}{(TP + FP)}$$
$$recall = \frac{TP}{(TP + FN)}$$
$$F_1\ measure = 2 \times \frac{(precision \times recall)}{(precision + recall)}$$

(iv) The **average class accuracy (harmonic mean)**. (During this calculation you should round all long floats to 3 places of decimal.)

(8 marks)

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

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	true	true
3	false	false	13	false	false
4	false	false	14	true	true
5	true	true	15	false	false
6	false	false	16	false	false
7	true	true	17	true	false
8	true	true	18	true	true
9	false	false	19	true	true
10	false	false	20	true	true

2. (a) A data analyst building a k-nearest neighbour model for a continuous prediction problem is considering appropriate values to use for k.

   (i) Initially the analyst uses a simple average of the target variables for the k nearest neighbours in order to make a new prediction. After experimenting with values for k in the range $0 - 10$ it occurs to the analyst that they might get very good results if they set k to the total number of instances in the training set. Do you think the analyst is likely to get good results using this value for k?

   (5 marks)

   (ii) If the analyst was using a distance weighted averaging function rather than a simple average for their predictions would this have made their idea any more useful?

   (5 marks)

   (b) A dataset showing the decisions made by an individual about whether to wait for a table at a restaurant is listed in Table 1 on the next page. (Note that Table 3, also on the next page, lists some equations that you may find useful for this question.)

   (i) Given that the WILLWAIT column lists the values of the target variable, compute the entropy for this dataset.

   (5 marks)

   (ii) What is the information gain for the PATRONS feature?

   (5 marks)

   (iii) What is the information gain for the TYPE feature?

   (5 marks)

   (iv) Given a choice between the PATRONS and TYPE feature, which feature would the ID3 algorithm choose as the root node for a decision tree?

   (5 marks)

ID	BAR	PATRONS	PRICE	RAIN	TYPE	WILLWAIT
1	F	Some	€€€	F	French	T
2	F	Full	€	F	Thai	F
3	T	Some	€	F	Burger	T
4	F	Full	€	F	Thai	T
5	F	Full	€€€	F	French	F
6	T	Some	€€	T	Italian	T
7	T	None	€	T	Burger	F
8	F	Some	€€	T	Thai	T
9	T	Full	€	T	Burger	F
10	T	Full	€€€	F	Italian	F
11	F	None	€	F	Thai	F
12	T	Full	€	F	Burger	T

Table 2: A dataset describing the previous decisions made by an individual about whether to wait for a table at a restaurant.

Table 3: Equations from information theory.

$$H(\mathbf{f}, \mathcal{D}) = - \sum_{l \in levels(f)} P(f = l) \times log_2(P(f = l))$$

$$rem(\mathbf{f}, \mathcal{D}) = \sum_{l \in levels(f)} \frac{|\mathcal{D}_{f=l}|}{|\mathcal{D}|} \times H(t, \mathcal{D})$$

$$IG(\mathbf{d}, \mathcal{D}) = H(\mathbf{t}, \mathcal{D}) - rem(\mathbf{d}, \mathcal{D})$$

Table 4: Query Document

Machine learning is fun

Table 5: Document counts from the corpus for the words in the query.

Document counts for the DISLIKE data set

fun	is	machine	learning
415	695	35	70

Document counts for the LIKE data set

fun	is	machine	learning
200	295	120	105

3. Lets assume we are given a set of **700** training documents that a friend has classified as DISLIKE and another **300** documents that they have classified as LIKE. We are now given a new document and asked to classify it. Table 4 lists the content of this query document and Table 5 gives the number of documents from each class (DISLIKE and LIKE) that the words in the query document occurred in. **What class will a Naive Bayes prediction model label the query document as belonging to?** (You must support your answer by showing the calculations that a Naive Bayes model will make.)

(30 marks)

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

$$\begin{aligned}\textsc{Repeat Purchase} = & -3.82398 - 0.02990 \times \textsc{Age} \\ & + 0.74572 \times \textsc{Shop Frequency} \\ & + 0.02999 \times \textsc{Shop Value}\end{aligned}$$

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 6, 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:

$$\begin{aligned}
&\phi_0(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = 1 && \phi_1(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = \textsc{Dose1} \\
&\phi_2(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = \textsc{Dose2} && \phi_3(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = \textsc{Dose1}^2 \\
&\phi_4(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = \textsc{Dose2}^2 && \phi_5(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = \textsc{Dose1}^3 \\
&\phi_6(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = \textsc{Dose2}^3 && \phi_7(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) = \textsc{Dose1} \times \textsc{Dose2}
\end{aligned}$$

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

$$\begin{aligned}
& P(\textsc{Type} = \textit{dangerous}) = \\
& Logistic\big(-0.848 \times \phi_0(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) + 1.545 \times \phi_1(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) \\
& \quad - 1.942 \times \phi_2(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) + 1.973 \times \phi_3(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) \\
& \quad + 2.495 \times \phi_4(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) + 0.104 \times \phi_5(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) \\
& \quad + 0.095 \times \phi_6(\langle \textsc{Dose1}, \textsc{Dose2} \rangle) + 3.009 \times \phi_7(\langle \textsc{Dose1}, \textsc{Dose2} \rangle)\big)
\end{aligned}$$

Use this model to make predictions for the query instances in Table 7 and using these prediction explain whether or not the dosage combinations are likely to lead to a dangerous of safe interaction.

(18 marks)

Table 6: 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 7: 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