S228/419C

**DUBLIN INSTITUTE OF TECHNOLOGY**
**KEVIN STREET, DUBLIN 8**

---

# BSc. (Honours) Degree in Computer Science

**Year 4**

---

## SUMMER EXAMINATIONS 2015

---

## ARTIFICIAL INTELLIGENCE II [CMPU4011]

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

Monday $11^{th}$ May 2015
4:00 p.m to 6:00 p.m

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

(5 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.)

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

(8 marks)

(iv) Which of these performance metrics (**misclassification rate** or **average class accuracy (harmonic mean)** is the most appropriate metric to use in this scenario? Provide an explanation for your answer.

(12 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) Table 3, on the next page, lists a dataset containing examples described by two descriptive features, **Feature 1** and **Feature 2**, and labelled with a target class **Target**. Table 4, also on the next page, lists the details of a query for which we want to predict the target label. We have decided to use a **3-Nearest Neighbor** model for this prediction and we will use Euclidean distance as our distance metric:

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^{n} \left( (x_1.f_i - x_2.f_i)^2 \right)}$$

(i) With which target class (**C1** or **C2**) will our **3-Nearest Neighbor** model label the query? Provide an explanation for your answer.

(8 marks)

(ii) There is a large variation in range between **Feature 1** and **Feature 2**. To account for this we decide to normalize the data. Compute the normalized versions of Feature 1 and Feature 2 to four decimal places of precision using range normalization

$$x_i.f' = \frac{x_i.f - min(f)}{max(f) - min(f)}$$

(4 marks)

(iii) Assuming we use the normalized dataset as input, with which target class (**C1** or **C2**) will our **3-Nearest Neighbor** model label the query? Provide an explanation for your answer.

(8 marks)

(b) Table 5, on the next page, lists a classification dataset. Each instance in the dataset has two descriptive features (Feature A and Feature B) and is classified as either a positive (+) or a negative(-) example. Note that Table 2, below, lists some equations that you may find useful for this question.

Table 2: Equations from information theory.

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

(i) Calculate the classification **entropy** for this dataset.

(5 marks)

(ii) Calculate the **information gain** for Feature A and Feature B.

(5 marks)

Table 3: Dataset for the 3-Nearest Neighbor question

| ID | Feature 1 | Feature 2 | Target |
|---|---|---|---|
| 101 | 4 | 180000 | C1 |
| 102 | 3 | 120000 | C2 |
| 103 | 7 | 360000 | C2 |
| 104 | 5 | 420000 | C1 |
| 105 | 8 | 480000 | C2 |

Table 4: Query instance for the 3-Nearest Neighbor question.

| ID | Feature 1 | Feature 2 | Target |
|---|---|---|---|
| 250 | 4 | 240000 | ? |

Table 5: Classification dataset for information question.

| Feature A | Feature B | Classification |
|---|---|---|
| True | True | + |
| True | False | - |
| True | False | + |
| True | True | + |
| False | True | - |

3. Table 6 lists a dataset of the previous decision made by a couple regarding whether or not they would wait for a table at a restaurant (i.e., the feature WAITED 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 (WAITED=Yes, and WAITED=No) for the following restaurant for this couple (marks will be deducted if workings are not shown, round your results to four places of decimal)

   BAR=False, PATRONS=None, PRICE=Expensive

   (10 marks)

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

   (2 marks)

Table 6: A dataset describing the previous decisions made by an individual about whether to wait for a table at a restaurant.

| ID | BAR | PATRONS | PRICE | WAITED |
|---|---|---|---|---|
| 1 | False | Some | Expensive | Yes |
| 2 | False | Full | Cheap | No |
| 3 | True | Some | Cheap | Yes |
| 4 | False | Full | Cheap | Yes |
| 5 | False | Full | Expensive | No |
| 6 | True | Some | Reasonable | No |
| 7 | True | None | Cheap | No |
| 8 | False | Some | Reasonable | No |
| 9 | True | Full | Cheap | No |
| 10 | True | None | Reasonable | Yes |

4. (a) The following model is commonly used for continuous prediction tasks:

$$y(x) = w_0 + w_1 x_1 + \cdots + w_D x_D$$

(i) Provide the name for this model and explain all of the terms that it contains.

(4 marks)

(ii) Explain how the following model can overcome some of the limitations of the model given above.

(8 marks)

$$y(x) = \sum_{j=0}^{M-1} w_j \phi_j(x)$$

(b) 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 7 below.

(18 marks)

Table 7: The queries for the multivariate logistic regression question

| ID | AGE | SHOP FREQUENCY | SHOP VALUE |
|---|---|---|---|
| 1 | 56 | 1.60 | 109.32 |
| 2 | 21 | 4.92 | 11.28 |
| 3 | 48 | 1.21 | 161.19 |