| PAPER CODE NO. | EXAMINER : Xiaowei Huang |
| --- | --- |
| **COMP219** | DEPARTMENT : Computer Science Tel. No. 0151 795 4260 |

UNIVERSITY OF LIVERPOOL

# First Semester Class Test 2018/19

# Advanced Artificial Intelligence

**TIME ALLOWED : 50 minutes**

---

**INSTRUCTIONS TO CANDIDATES**

Answer FOUR questions.

Please select correct answers. You need to answer all the questions. Questions may be assigned with different marks.

Intelligence

| Grade | Low | Average | High |
|---|---|---|---|
| A | 0.04 | 0.3 | 0.1 |
| B | 0.06 | 0.3 | 0.06 |
| C | 0.1 | 0.3 | 0.04 |

Figure 1: Joint probability for student grade and intelligence

**Question 1: Basic Knowledge (Total 25 marks)**

1. Which learning task best suits the following description: given a set of training instances $\{(x^{(1)}, y^{(1)}), ..., (x^{(n)}, y^{(n)})\}$ of an unknown target function $f$, where $x^{(i)}$ is the feature vector and $y^{(i)}$ is the label for $i \in \{1, ..., n\}$, it outputs a model $h$ that best approximates $f$. **2 marks**

   **(a)** Unsupervised learning

   **(b)** Supervised learning

   **(c)** Reinforcement learning

   **(d)** none of the above

2. Compute the following probability according to the table in Figure 1

$$P(Intelligence = High) =$$

**2 marks**

   **(a)** 0.28

   **(b)** 0.3

   **(c)** 0.35

   **(d)** 0.2

3. Compute the following conditional probability according to the table in Figure 1

$$P(Intelligence = High \mid Grade = C) =$$

**2 marks**

   **(a)** 0.04/0.44

   **(b)** 0.07/0.25

   **(c)** 0.35/0.7

   **(d)** 0.28/0.7

4. Which of the following statements are correct regarding to Figure 2. **3 marks**

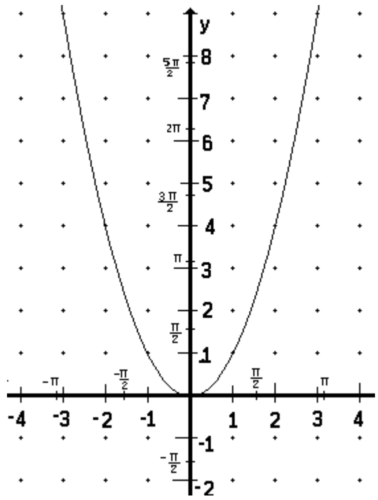


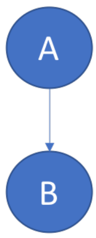

Figure 2: Diagram for $y = x^2$ function

A: Diseases

P(A)

| $a^0$ | $a^1$ |
|---|---|
| 0.3 | 0.7 |

B: Symptom




P(B|A)

| P(B\|A) | $b^0$ | $b^1$ |
|---|---|---|
| $a^0$ | 0.2 | 0.8 |
| $a^1$ | 0.4 | 0.6 |

Figure 3: Probabilistic Graph of Diseases and Sympton

**(a)** $\min_x x^2 = 1$

**(b)** $\min_x x^2 = 0$

**(c)** $0$ is in $\arg\min_x x^2$

**(d)** $\pi$ is in $\arg\min_x x^2$

**5.** Use the information provided in Figure 3 to compute the following joint probability

$$P(A = a^1, B = b^1) =$$

**2 marks**

**(a)** 0.06

**(b)** 0.36

**(c)** 0.24

**(d)** 0.42

**6.** Use the information provided in Figure 3 to compute the following expression

$$\max_{A,B} P(A, B) =$$

**2 marks**

**(a)** 0.42

**(b)** $a^1, b^1$

**(c)** 0.28

**(d)** $a^0, b^1$

**7.** Use the information provided in Figure 3 to compute the following maximum a posterior expression

$$MAP(A, B) =$$

**3 marks**

**(a)** 0.36

**(b)** $a^1, b^1$

**(c)** 0.5

**(d)** $a^0, b^1$

**8.** Understanding simple numpy command.
Assume that $a = np.arange(10).reshape((2, 5))$. Then $a.T.shape =$ **2 marks**

**(a)** 10

**(b)** (2,5)

**(c)** (5,2)

**(d)** 2

**9.** Let $x = (-2, 0, 6, -4)$ be a vector. Then its $L^2$ norm $||x||_2 =$ **2 marks**

**(a)** $\sqrt{30}$

**(b)** $\sqrt{56}$

**(c)** 12

**(d)** 2

**10.** Let $x = (0, 2, 3, -4)$ be a vector. Then its $L^1$ norm $||x||_1 =$ **3 marks**

**(a)** 2

**(b)** 9

**(c)** 4

**(d)** 3

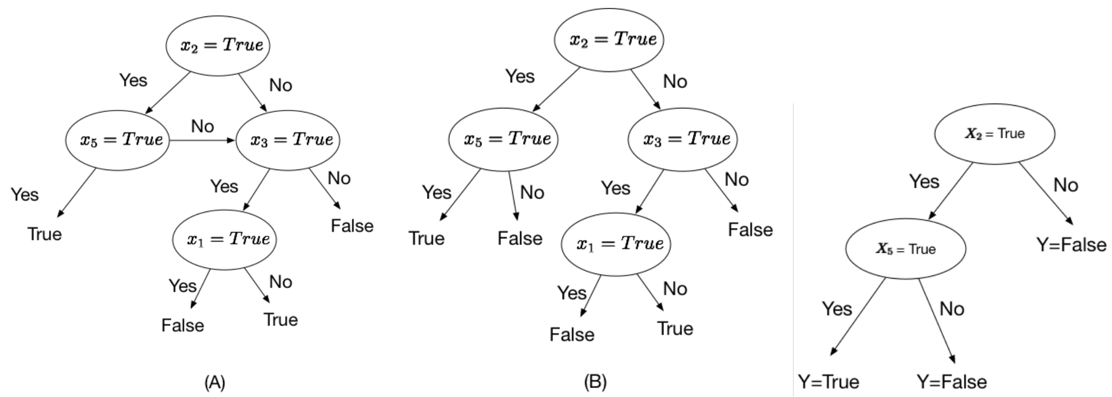



Figure 4: Decision Trees

| v1 | v2 | v3 | y |
|---|---|---|---|
| 182 | 87 | 11.3 | No |
| 189 | 92 | 12.3 | Yes |
| 178 | 79 | 10.6 | Yes |
| 183 | 90 | 12.7 | No |

Figure 5: Dataset for football payer

**Question 2: Simple Learning Models (Total: 33 marks)**

**1.** Which decision trees in Figure 4 can represent the boolean formula $(x_2 \wedge x_5)$ **3 marks**

- **(a)** A
- **(b)** B
- **(c)** C
- **(d)** none of the above

**2.** Figure 5 gives an example dataset $D$ about football player. Please indicate which of the following expressions is to compute its entropy $H_D(Y)$, where $Y$ is the random variable for labelling: **3 marks**

- **(a)** $-\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2})$
- **(b)** $-\frac{1}{2}\log_2(\frac{1}{2})$
- **(c)** $\frac{1}{2}\log_2(\frac{1}{2})$
- **(d)** none of the above

**3.** Figure 5 gives an example dataset $D$ about football player. Please compute the information gain of splitting over the feature Wind **InfoGain**$(D, V_1) = H_D(Y) - H_D(Y \mid V_1)$: **3 marks**

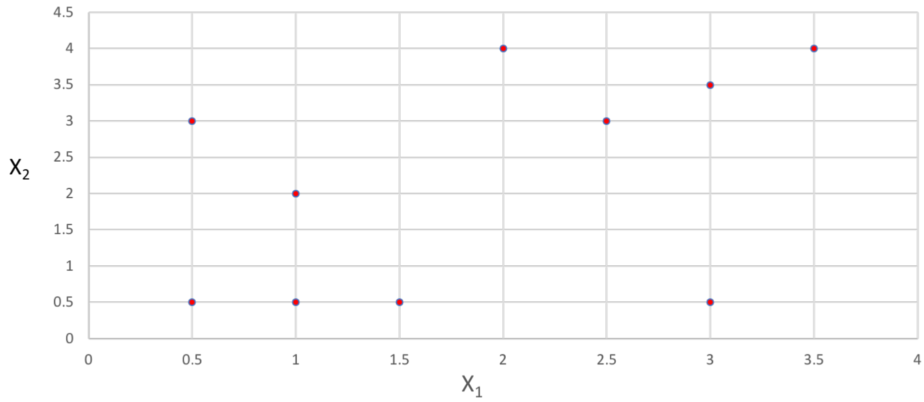



Figure 6: A set of two-dimensional input samples

**(a)**

**(b)**

**(c)**

**(d)**

**4.** Assume that, as shown in Figure 6, we have a set of training instances with two features $X_1$ and $X_2$:

$$\{(0.5, 3), (0.5, 0.5), (1, 0.5), (1, 2), (1.5, 0.5), (2, 4), (2.5, 3), (3, 0.5), (3, 3.5), (3.5, 4)\}$$

such that

- the instance $(0.5, 3), (1, 2), (2, 4), (2.5, 3), (3, 3.5), (3.5, 4)$ are labeled with value $1$,
- the other points are labelled with value 2

Now, we have a new input $(2, 2.5)$, please indicate which of the following points are *not* considered for the 3-nn classification, according to the manhattan ($L^1$) distance. **3 marks**

**(a)** (1,2)

**(b)** (1.5,0.5)

**(c)** (2.5,3)

**(d)** (2,4)

**5.** Continue with the above. Now, for new input $(2, 2.5)$, please compute its regression result for the $3$-nn regression, according to the Manhattan distance. **3 marks**

**(a)**

**(b)**

**(c)**

| | | Actual Class | |
|---|---|---|---|
| | | positive | negative |
| Predicted Class | positive | A | B |
| | negative | C | D |

Figure 7: A confusion matrix for two-class problem

**(d)**

**6.** Assume a two-class problem where each instance is classified as either 1 (positive) or -1 (negative). We have a training dataset of 1,000 instances, such that 600 of them are labeled as 1 and 400 of them are labeled as -1. After training, we apply the trained model to classify the 1,000 instances and find that 800 instances are classified the same as their true labels. Moreover, we know that, 500 instances are classified as 1 and, within the 500 instances, 50 instances are actually labeled as -1. Please indicate which numbers should be filled in to $(A, B, C, D)$ in Figure 7. **3 marks**

**(a)**

**(b)**

**(c)**

**(d)**

**7.** Continue with the above question. Please compute the error rate of the trained model. **3 marks**

**(a)**

**(b)**

**(c)**

**(d)**

**8.** Given training data $\{(x^{(i)}, y^{(i)}) \mid 1 \leq i \leq m\}$, which one of the following is closest to the logistic regression **3 marks**

**(a)** minimise the loss $\hat{L} = \dfrac{1}{m}\sum_{i=1}^{m}(w^T x^{(i)})$

**(b)** minimise the loss $\hat{L} = \dfrac{1}{m}\sum_{i=1}^{m}(\sigma(w^T x^{(i)}) - y^{(i)})^2$, where $\sigma(a) = \dfrac{1}{1 + exp(-a)}$

**(c)** minimises the loss $\hat{L} = \dfrac{1}{m}\sum_{i=1}^{m}(w^T x^{(i)} - y^{(i)})^2$

**(d)** minimise the loss $\hat{L} = \frac{1}{m}\sum_{i=1}^{m}(\log(w^T x^{(i)}) - y^{(i)})^2$

**9.** Let $f(X) = 3X_1^2 + 4X_2^8 + 5X_3$ be a function, where $X_1$, $X_2$ and $X_3$ are three variables. Please indicate which of the following gradient expression is correct: **3 marks**

**(a)** $\nabla_X f(X) = 6X_1$

**(b)** $\nabla_X f(X) = (3, 4, 5)$

**(c)** $\nabla_X f(X) = (6X_1, 28X_2, 5)$

**(d)** $\nabla_X f(X) = (6X_1, 28X_2^8, 5)$

**10.** Naive Bayes method is based on the following assumption, where $X_i$ for $i \in \{1..n\}$ represent features of an instance and $Y$ represents the parameter: **3 marks**

**(a)** $P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i)$

**(b)** $P(X_1, ..., X_n \mid Y) = \prod_{i=1}^{n} P(X_i \mid Y)$

**(c)** $P(X_1, ..., X_n) = \sum_{i=1}^{n} P(X_i)$

**(d)** $P(X_1, ..., X_n \mid Y) = \sum_{i=1}^{n} P(X_i \mid Y)$

This page collects some formulas/expressions that may be used in this exam.

**1.** entropy:

$$-\sum_{y \in values(Y)} P(y) \log_2 P(y)$$

**2.** conditional entropy:

$$H(Y|X) = \sum_{x \in values(X)} P(X = x) H(Y|X = x)$$

where

$$H(Y|X = x) = -\sum_{y \in values(Y)} P(Y = y|X = x) \log_2 P(Y = y|X = x)$$