Q1 Logical Agents

21 Points

Justify all answers

Consider the following knowledge base:

- If there is a bug, then the code will not execute.
- If there is a syntax error or an infinite loop, then there is a bug in the code.
- The program did execute.

Q1.1

7 Points

Translate these premises into propositional logic using the connectives NOT, AND, OR and IMPLIES. Use only the following atomic propositions: $\mathbf{loop}, \mathbf{bug}, \mathbf{execute}, \mathbf{syntax}$.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

No files uploaded

Q1.2

7 Points

Translate the propositions into the conjunctive normal form.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

No files uploaded

Q1.3
7 Points

Use the resolution algorithm to prove "$\neg syntax$" . Show all your work. One resolution per line. You can use NOT for the negation, and mark the contradiction with the word CONTRADICTION.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

No files uploaded

Q2
27 Points

Justify all answers

Consider the following one-dimensional dataset. Each example is described by one numerical feature x and has a discrete label $y \in \{0, 1\}$.

x	1	2	3	6	6	7	10	11
y	0	0	1	1	1	0	0	0

Q2.1
9 Points

Consider $x \in [0, 12]$. What is the output of a 1-NN (nearest neighbor classifier)? Provide ALL the intervals in [0, 12] labeled 1. For example, [6, 7] is labeled 1. In case of ties, choose the point on the left.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

2.1.png Download

The interval is $2.5 < x \leq 6.5$, where the output is ONE. That is $(2.5, 6.5] = 1$

Q2.2

9 Points

Consider $x \in [0, 12]$. What is the output of a 5-NN (nearest neighbor classifier)?

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

2.2.png Download

We have $x \in [0, 7]$, with the classifier predicting ONE. And have $x \in (7, 12]$, with the classifier predicting ZERO. For any inputs in the first interval, there are always three nearest neighbours with labels of ONE, whereas for those inputs in the second interval, there are always three nearest neighbours with labels of ZERO.

Q2.3

9 Points

Assume we do an 8-fold cross-validation (hint: one point per fold). Using 1-NN, what would be the cross-validation accuracy? Show your work.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

2.3.png Download

We have $k = 1$:

1. $i = 1$, then the first data point $(x = 1, y = 0)$ is labelled CORRECTLY according to the other seven data points. (The current NN is $x = 2, y = 0$)
2. $i = 2$, then the second data point $(x = 2, y = 0)$ is labelled CORRECTLY according to the rest seven data points. (The current NN is $x = 1, y = 0$)
3. $i = 3$, then the third data point $(x = 3, y = 1)$ is labelled INCORRECTLY according to the rest seven data points. (The current NN is $x = 2, y = 0$)
4. $i = 4$, then the fourth data point $(x = 6, y = 1)$ is labelled CORRECTLY according to the rest seven data points. (The current NN is $x = 6, y = 1$)
5. $i = 5$, then the fifth data point $(x = 6, y = 1)$ is labelled CORRECTLY according to the rest seven data points. (The current NN is $x = 6, y = 1$)
6. $i = 6$, then the sixth data point $(x = 7, y = 0)$ is labelled INCORRECTLY according to the rest seven data points. (The current NN is $x = 6, y = 1$)
7. $i = 7$, then the seventh data point $(x = 10, y = 0)$ is labelled CORRECTLY according to the rest seven data points. (The current NN is $x = 11, y = 0$)
8. $i = 8$, then the eighth data point $(x = 11, y = 0)$ is labelled CORRECTLY according to the rest seven data points. (The current NN is $x = 10, y = 0$)

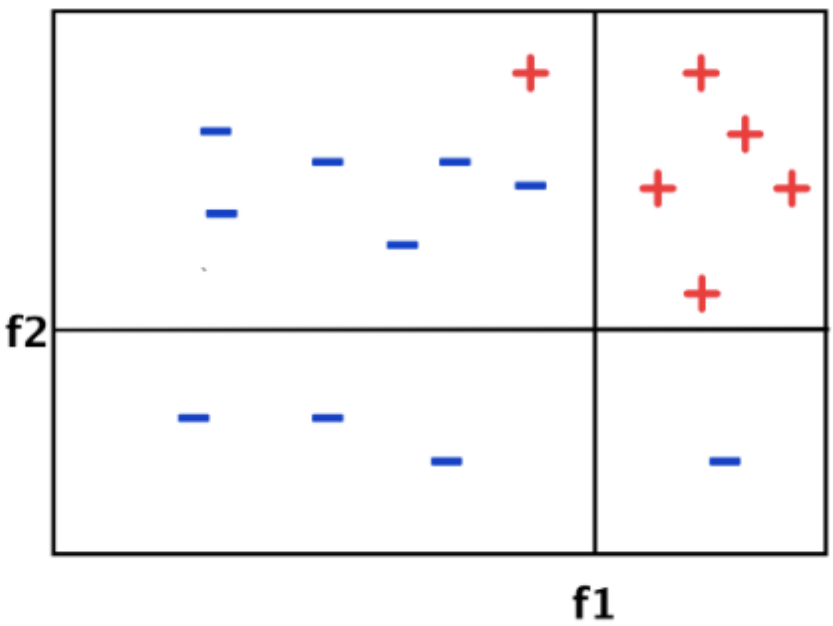
In summary, our accuracy is $\frac{6}{8} = 75\%$

Q3 Decision Trees

25 Points

Justify all answers

Which of the features f_1 or f_2 will be chosen as the initial split by the decision tree algorithm? Justify your work using Entropy and Information gain calculations.



Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

▾ 3.png Download

For the initial state, we have

$$\text{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2 p_i = -\frac{6}{16}\log_2(\frac{6}{16}) - \frac{10}{16}\log_2(\frac{10}{16}) \approx 0.954 \quad (5)$$

If we use f_1 as the first node, we have the information gain

$$\begin{aligned}\text{Gain}(S, A) &= \text{Entropy}(S) - \sum_{v \in \text{ Values (A)}} \frac{|S_v|}{|S|} \text{Entropy}\left(S_v\right) \\ &\approx 0.954 - \frac{10}{16}(-\frac{1}{10}\log_2\frac{1}{10} - \frac{9}{10}\log_2\frac{9}{10}) + \frac{6}{16}(-\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6}) \\ &\approx 0.954 - 0.699 = 0.255.\end{aligned}$$

If we use f_2 as the first node, we have the information gain

$$\begin{aligned}\text{Gain}(S, A) &= \text{Entropy}(S) - \sum_{v \in \text{ Values (A)}} \frac{|S_v|}{|S|} \text{Entropy}\left(S_v\right) \\ &\approx 0.954 - \frac{4}{16}(-\frac{0}{4}\log_2\frac{0}{4} - \frac{4}{4}\log_2\frac{4}{4}) + \frac{12}{16}(-\frac{5}{12}\log_2\frac{5}{12} - \frac{7}{12}\log_2\frac{7}{12}) \\ &\approx 0.954 - 0.735 = 0.219,\end{aligned}$$

where we used the fact that $\lim_{x\to 0}(x\log_2 x) = 0$

According to information gained from choosing f_1 and f_2 as the initial split, we know that f_1 has a better behaviour since information gain for f_1 is approximately 0.255, whereas that for f_2 is approximately 0.219. Therefore, f_1 will be chosen as the initial split rule.

Q4 Linear Regression

27 Points

Justify all answers

Consider the problem of predicting student scores in the final exam in function of their scores in the midterm exam. Let feature x denote the midterm scores and label y denote the final scores. Consider the training set pairs (x, y) as follows: (55, 67), (60, 63), (66, 72), (72, 90), (85, 93), (90, 92).

Q4.1

5 Points

Suppose we learned a linear classifier and the weights are $\beta_0 = -8$ and $\beta_1 = 1.2$. what is the predicted final score if the midterm grade is 80?

Show your work.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

▾ 4.1.png Download

We have

$$y = \beta_0 + \beta_1 x = -8 + 1.2x. \tag{5}$$

If $x = 80$, then the prediction of the final score, y, will be 88.

Q4.2

10 Points

Recall the cost function R seen in class that depends on the regression weights. Calculate the cost for $\beta_0 = -8$ and $\beta_1 = 1.2$.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

▾ 4.2.png Download

Our data set is $\{(55, 67), (60, 63), (66, 72), (72, 90), (85, 93), (90, 92)\}$. Thus $n = 6$. We have $f(x) = -8 + 1.2x$. So

$$\begin{aligned} R =& \frac{1}{2n}\sum_{i=1}^{n}(y_i - f(x_i))^2 \\ =& \frac{1}{12}((67-58)^2 + (63-64)^2 + (72-71.2)^2) + (90-78.4)^2 + (93-94)^2 + (92-100)^2) \\ =& \frac{282.2}{12} \approx 23.52. \end{aligned}$$

That is, our cost over the training set is approximately 23.52

Q4.3

12 Points

Suppose we managed to train a linear regression on the training data and we found β_0 and β_1 such that $R = 0$. Which of the following is

correct/incorrect. Explain each answer.

(a) We must have β0 = 0 and β1 = 0.
(b) We have found a linear regressor that perfectly fit the data.
(c) We will do a perfect prediction in the test set.

Only do one of the answer options, a text entry or a screenshot of latex, not both. (If you do both we will grade only the text entry.)

4.3.png Download

1. (a) is INCORRECT. If our training data is $\{(1, 1), (2, 2), (3, 3)\}$, we may find $\beta_0 = 0$ but $\beta_1 = 1$ instead. And we still have $R = \frac{1}{2n}\sum_{i=1}^{n}(y_i - f(x_i))^2$ is still ZERO, where $n = 3$ for our $f(x) = x$.
2. (b) is CORRECT. If we have $R = 0$, this guarantees that in the equation $R = \frac{1}{2n}\sum_{i=1}^{n}(y_i - f(x_i))^2$, we have for each data point, (x_i, y_i), $y_i = f(x_i)$. That is, we perfectly fit the data.
3. (c) is INCORRECT. We train our $f(x)$ over the domain of the training data only, and the real distribution of the data remains unclear. For the example in (a), we may have a test data point of $(100, 99)$, which can not be perfectly predicted by our $f(x) = x$. More general comments: in practice, when $R = 0$, we may overfit the training data, which can introduce a higher error in the test set.

Q5 Deadline

0 Points

Please make a note here if you received an extension from the teaching staff. If not, please leave this blank. We will use this question for early/late submission assignment adjustments.

No extension received

HW3 Conceptual

● UNGRADED

STUDENT

Ziggy Chen

TOTAL POINTS

- / 100 pts

QUESTION 1

Logical Agents — 21 pts

1.1 (no title) — 7 pts

1.2 (no title) — 7 pts

1.3 (no title) — 7 pts

QUESTION 2

(no title) — 27 pts

2.1 (no title) — 9 pts

2.2 (no title) — 9 pts

2.3 (no title) — 9 pts

QUESTION 3

Decision Trees — 25 pts

QUESTION 4

Linear Regression — 27 pts

4.1 (no title) — 5 pts

4.2 (no title) — 10 pts

4.3 (no title) — 12 pts

QUESTION 5

Deadline — 0 pts