Machine Learning Task 1

The dataset and instructions for this task are available in the pdf released by Coding Club.
Attempt the task honestly. All the best!

Note:
1. There is no negative marking, so attempt all questions.
2. For questions with multiple right answers, there is no partial marking.

1

Name *

Sri Kanishk S

2

Roll Number *

200102091

3

IITG email ID *

kanishk@iitg.ac.in

4

Among the different categories of students who consume alcohol on working days, which category has the lowest average G3 score?
(2 Points)

- 1
- 2
- 3
- 4
- 5

5

Suppose that scientists have discovered a new ultra-fast algorithm for matrix multiplication for large matrices. Which of the below two is more likely to be benefitted from this discovery?
(2 Points)

- Linear Regression
- Logistic Regression

6

Find the following information about the topper(the student with the highest G3 score):
i. Sex
ii. School
iii. Relationship Status
iv. Absences
(2 Points)

- Male, GP, single, 4
- Female, GP, single, 3
- Male, MS, single, 4

Male, MS, not single, 5

7

Let f be some arbitrary smooth function of two variables x and y. Suppose we use gradient descent to try to minimize f(x,y) as a function of x and y. Which of the following are true?

Note: multiple answers may be correct. Select all that are true.
(4 Points)

- [ ] If x and y are initialized so that x=y, then after one iteration of gradient descent, we will still have x=y.
- [ ] Every iteration of gradient descent will decrease the value of f(x,y) irrespective of the value of the learning rate α.
- [ ] As long as learning rate is small enough, gradient descent will converge to the same solution regardless of how x and y are initialized.
- [x] If x and y are initialized at a local minimum, then one iteration will not change their values.
- [x] If the learning rate is too small, then gradient descent may take a very long time to converge.

8

Suppose you need a G3 score >= 15 in order to be eligible for higher education.
Among those students who wish to pursure higher education, find what percentage of them are eligible.
Round your answer off to 2 decimal places.
(2 Points)

73

9

Find the absolute difference between the average grades of the students who are in romantic relationship and those who are not in a romantic relation.
Round your answer off to 2 decimal places.
(2 Points)

0.8

10

Suppose that for some linear regression problem we have some training set, and for our training set, we managed to find some $\theta 0$, $\theta 1$ such that $J(\theta 0,\theta 1)=0$.

Consider the following statements.

1. For this to be true, we must have $y(i)=0$ for every value of $i=1,2,\ldots,m$.

2. Gradient descent is likely to get stuck at a local minimum and fail to find the global minimum.

3. For this to be true, we must have $\theta 0=0$ and $\theta 1=0$ so that $h\theta(x)=0$

4. Our training set can be fit perfectly by a straight line.

5. We can perfectly predict the value of y even for new examples that we have not yet seen.

How many of the above statements are true?
(4 Points)

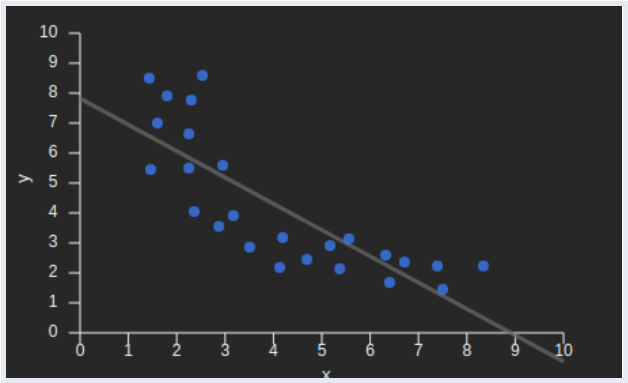
1

11

Suppose we define a new quantity TotalAlc, where TotalAlc = Walc + Dalc. Find the Pearson correlation between TotalAlc column and the famrel column.
Round your answer off to 2 decimal places.
(2 Points)

-0.11

12



Suppose we run linear regression on the data given in the graph below and end up with the best-fitting line(the gray line). Suppose the line represents the equation y = mx + c.
Which of the following are the MOST PLAUSIBLE values of the parameters m and c?
(2 Points)

- m = 0.9, c = 8
- m = 0.9, c = -8
- m = -0.9, c = 8
- m = -0.9, c = -8

13

Consider the following data, where x(the independent variable) represents the marks obtained in an exam, and y(the dependent variable) represents the scholarship amount offered(in lakhs). Assuming we run linear regression until convergence on this data, find the scholarship amount(in lakhs) obtained by a student who scored 1 mark.

x	y
2.50	4
1.75	2.5
0.50	0
3.00	5

(2 Points)

1

14

Which sex has more percentage participation in extra curricular activities? Also specify the percentage.
(2 Points)

- M, 59.2%
- F, 56.1%
- M, 56.1%
- F, 59.2%

15

How many of the following statements are false? Note that performance is measured by G3 scores.

"On an average, boys performed better than girls."
"On an average, boys studied for more time than girls."
"On an average, boys from rural areas outperformed girls from urban areas."
(2 Points)

- 0
- 1
- 2
- 3