Pravin Patil → End sem- Part 2

उत्तर पुस्तिका/ Answer Book

भारतीय प्रौद्योगिकी संस्थान मुंबई
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

X2ptp512
रोल नं./Roll No.

P.M.L.
पाठ्यक्रम नाम/Course Name

Aug
शाखा/प्रभाग/Branch/Div. शैक्षणिक बैच /Tutorial Batch

अनुभाग/Section

Q.1

(a)

a ~~(a)~~ = [0, 1, 2, 3, 4] → array with length 5

f(a) = f(list[0, 1, 2, 3, 4], it appends last index value

f(a) = [0, 1, 2, 3, 4, 16]

~~a = 21~~ X = [0, 1, 2, 3, 4, 16]

(b) lambda → is inline function

It ~~return false~~ true if number is odd

f(11)

f = 11 % 2 == 1

f(11) → true

Pravin Patil Part 2 Page 2

Q.1

(c)

r = f(3)

y = 3//2 + 1 = 1 + 1 = 2

y, y*y => 2, 3*2 = 6

→ r = (2, 6) → tuple

since r[0] = 3.3/2

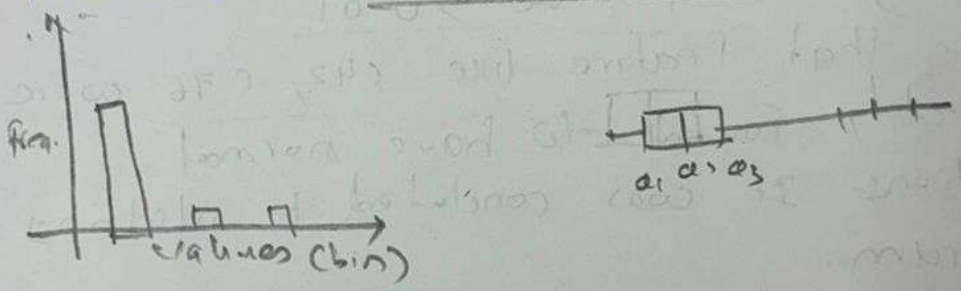
~~r ==~~ (function returned tupled

print(r) → might give compilation error.

Q.2

(9) The box plots and histograms helps to understand skewness. In the given excersie, we plot box plot & histogram of both skewed and non skewed feature

Feature like C123, has higher skewness, so box plot was wider indicating there were outliers while histogram also showed that there were more data points on the left



(3) We take example of C123 where skewness was ~ 38.10 which is pretty high.
Box cox transformation & log transformation helped to reduce the skewness.

Pravin Pathi part 2

Q.1 → part(d)

There are 3 functions defined here.

Function 1 :- update

parameters :- (guess) - which could be integer

return value: It returns addition of reciprocal of guess and adds 1

Example update (10) => $1/10+1 = 1.1$

Function 2 :- close

parameters :- (g) - Integer

Returns :- boolean if absolute value of $(g^2 - (g+1))$ is less than 0.1

Example :- close (10) = ? $(10^2 - (10+1)) < 0.1$

= false

Function 3 improve -

parameters :-

1. function update
2. function close

for close there is default value for parameter guess

improve(update, close = close)

→ guess = 1

close (1) => (False)

guess = update (1) => $\frac{1}{1} + 1 = 2$

→ 2 is ~~$\left(\frac{2}{1+2}\right)+3$~~ = $\left(\frac{2}{2}\right)+1 = 2$

guess = 2

close (2) $[4-(3)] < 0.1$

$1 < 0.1$

guess = update (2) => $2/2 + 1 = 2$

$(0.5+1)$

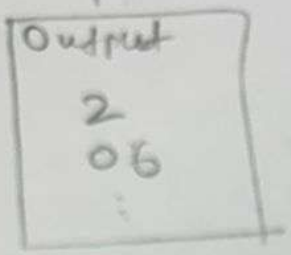
$(0.5 < 0.1)$

$(0.5)(-(0.7))$

$= .024 < 1$

Continued (PTO

Pravn Rathi part 2 pose 4



in any of the exercises) be sure to mention and explain them.

- **(a)** Is it correct to say that Box Plots and Histograms are closely related to each other? ... the help of concrete evid... on the provided dataset. **[2 marks]**

...aset, illustrate how data sh...

pg - 6

Q.2 (c) Normal distribution

We tried to plot ~~data~~ distributions of all the features using normal parameters like kurtosis and pvalue

Initial observation:-
with pvalue > 0.05 → There were no distinct feature which falls under normal distribution

So, Next step, is we checked the features which are more likely to be normally distributed by checking the pvalue > 0.01

So, After that feature like c42, c76 were approximately found to have normal distributions. It was concluded by plotting histogram.

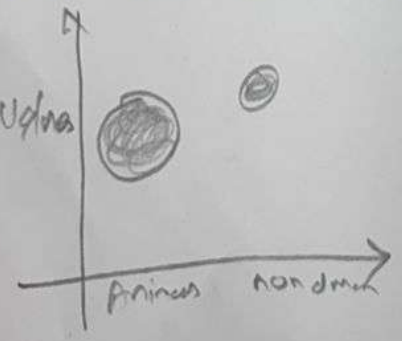
Q.2 (d)

t-SNE is used for plotting.

Based on the scores like Silhoit and Bowid

a) Silhot score (lower the better), cluster number = 2, after that silhoud score was dropped significantly.

optimal number of cluster = 2



1. Asked
   Bc
2. Com
3. He
   th
4. pl

pg - 7

(e)

1. Asked LLM to generate Silhot score & Bcmoid score for imputed datasets.
2. Compare both models and print accuracy.
3. Help to write conclusion based on the both the scores.
4. plot these clusters and name them.