

TRƯỜNG ĐẠI HỌC VĂN LANG  
KHOA CÔNG NGHỆ THÔNG TIN

**ĐỀ THI VÀ ĐÁP ÁN**  
**THI KẾT THÚC HỌC PHẦN**  
**Học kỳ 1, năm học 2023-2024**

**I. Thông tin chung**

Tên học phần:	Nhập môn Phân tích Dữ liệu và Học sâu		
Mã học phần:	DIT0550, 71ITDS30203	Số tín chỉ:	3
Mã nhóm lớp học phần:	231_DIT0550_01, 231_71ITDS30203_02		
Hình thức thi: <b>Trắc nghiệm kết hợp Tự luận</b>	Thời gian làm bài:	<b>90</b>	phút
<i>Thí sinh được tham khảo tài liệu:</i>	<input checked="" type="checkbox"/> Có	<input type="checkbox"/> Không	

**II. Các yêu cầu của đề thi nhằm đáp ứng CLO**

Ký hiệu CLO	Nội dung CLO	Hình thức đánh giá	Trọng số CLO trong thành phần đánh giá (%)	Câu hỏi thi số	Điểm số tối đa	Lấy dữ liệu đo lường mức đạt PLO/PI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
<b>CLO1</b>	Hiểu được nền tảng phân tích dữ liệu dựa trên các thuật toán, quy trình phát triển ứng dụng (thư viện phân mềm) dùng trong phân tích dữ liệu bằng ngôn ngữ lập trình Python	Trắc nghiệm	43.4%	1- 31	4.34	PLO1/P I1.2
<b>CLO2</b>	Làm quen với các khái niệm cơ bản của Học sâu, hiểu biết về các thư viện và công cụ sử dụng phổ biến trong Học sâu.	Trắc nghiệm+ Tự luận	56.6%	32-50 + Câu tự luận	5.66	PLO1/P I1.2

**III. Nội dung câu hỏi thi**

**PHẦN TRẮC NGHIỆM ( 50 câu (7 điểm))**

**Cấu trúc dữ liệu chính của Pandas là?**

- A. DataFrame
- B. Keyframe
- C. Statistics
- D. Econometrics

ANSWER: A

**Một Panel là một container \_\_\_\_ của dữ liệu**

- A. 3D
- B. 1D
- C. 2D
- D. Infinite

ANSWER: A

**Đoạn code sau đây xuất ra kết quả gì?**

```
import pandas as pd
s = pd.Series([1,2,3,4,5],index = ['a','b','c','d','e'])
print s['a']
```

- A. 1
- B. 2
- C. 3
- D. 4

ANSWER: A

**Cú pháp tạo DataFrame trong pandas?**

- A. pandas.DataFrame( data, index, columns, dtype, copy)
- B. pandas.DataFrame( data, index, dtype, copy)
- C. pandas.DataFrame( data, index, rows, dtype, copy)
- D. pandas\_DataFrame( data, index, columns, dtype, copy)

ANSWER: A

**\_\_\_\_\_ là một thư viện quan trọng được sử dụng trong phân tích dữ liệu.**

- A. Pandas
- B. Math
- C. Random
- D. Không có câu nào

ANSWER: A

**Thư viện nào sau đây trong Python được sử dụng để vẽ các đồ thị và trực quan hóa?**

- A. Matplotlib
- B. Pandas
- C. NumPy
- D. Không có câu nào

ANSWER: A

**Câu lệnh nào sau đây được sử dụng để cài đặt thư viện pandas?**

- A. pip install pandas

- B. install pandas
  - C. pip pandas
  - D. Không có câu nào
- ANSWER: A

**Một \_\_\_\_\_ là một mảng 1 chiều.**

- A. Series
  - B. Data Frame
  - C. Cả 2 câu trên
  - D. Không có câu nào
- ANSWER: A

**Nhãn dữ liệu được liên kết với một giá trị cụ thể của chuỗi được gọi là \_\_\_\_\_**

- A. Index
  - B. Data value
  - C. Value
  - D. Không có câu nào
- ANSWER: A

**Viết kết quả xuất ra của câu lệnh sau:**

```
>>> S1=pd.Series([14, 7, 9] ,index = range(1, 8, 3))
>>> print(S1)
```

**A.**  
1 14  
4 7  
7 9  
dtype: int64

**B.**  
14 1  
7 4  
9 7  
dtype: int64

- C. Error
  - D. Không có câu nào
- ANSWER: A

**Hàm nào sau đây cung cấp thông tin về dữ liệu ở mức cao nhất?**

- A. head
  - B. tail
  - C. summary
  - D. Không có câu nào
- ANSWER: A

**Theo mặc định, các giá trị bị thiếu trong bất kỳ tập dữ liệu nào được đọc là \_\_\_\_\_**

- A. NaN
- B. NA
- C. NULL

D. 0

ANSWER: A

**Phương thức nào sau đây được sử dụng để điền các giá trị null bằng một giá trị mặc định?**

A. fillna

B. fill

C. keepna

D. keep

ANSWER: A

**Phương thức nào sau đây của pandas được sử dụng để kiểm tra xem mỗi giá trị có phải là giá trị rỗng hay không?**

A. isnull

B. NULL

C. isnan

D. ifnull

ANSWER: A

**Phương thức nào sau đây được sử dụng để loại bỏ các dòng trùng nhau?**

A. drop\_duplicates

B. remove\_dup

C. remove

D. drop\_dup

ANSWER: A

**Phương thức nào sau đây được sử dụng để loại bỏ các dòng có giá trị null?**

A. dropna

B. drop

C. remove

D. removena

ANSWER: A

**Phương thức nào sau đây được sử dụng để nhóm dữ liệu của DataFrame dựa trên các cột cụ thể?**

A. groupby

B. aggregate

C. group

D. groupat

ANSWER: A

**Giả sử có một DataFrame tên là df với các cột ['A', 'B', 'C', 'D'] và các dòng ['r1', 'r2', 'r3']. Câu lệnh nào sau đây lọc các dòng có giá trị cột B lớn hơn 45 và giá trị cột 'C' nhỏ hơn 30?**

A. df.loc[(df.B > 45) & (df.C < 30)]

B. df[df.B > 45 & df.C < 30]

C. df.loc[df.B > 45 & df.C < 30]

D. (df.B > 45) & (df.C < 30)

ANSWER: A

**Giả sử có một DataFrame tên là df với 10 dòng và chỉ mục index ['r1', 'r2', 'r3', 'row4', 'row5', 'row6', 'r7', 'r8', 'r9', 'row10']. Phương thức tổng hợp (aggregate method) ở đoạn mã dưới đây dùng để làm gì?**

```
g = df.groupby(df.index.str.len())
```

```
g.aggregate({'A':len, 'B':np.sum})
```

- A. Tính độ dài của cột A và tổng giá trị của cột B của mỗi nhóm
- B. Tính tổng các giá trị của cột A
- C. Tính độ dài của cột A
- D. Tính độ dài của cột A và tổng giá trị của cột B

ANSWER: A

**Phương thức nào sau đây được sử dụng để nối hai hoặc nhiều DataFrame?**

- A. concat
- B. con
- C. concatenate
- D. truncate

ANSWER: A

**Viết đoạn mã để sắp xếp DataFrame đã cho (giả sử DataFrame tên là df và bao gồm thông tin chi tiết của sinh viên) theo thứ tự giảm dần của AdmNo.**

- A. df.sort\_values('AdmNo', ascending=False)
- B. df.sort\_index('AdmNo', ascending=False)
- C. df.sort\_values('AdmNo', ascending=True)
- D. df.sort\_values('AdmNo', ascending=false)

ANSWER: A

**Gói python nào được sử dụng cho đồ họa 2D?**

- A. matplotlib.pyplot
- B. matplotlib.pip
- C. matplotlib.numpy
- D. matplotlib.plt

ANSWER: A

**Để vẽ biểu đồ x so với y, chúng ta có thể dùng lệnh \_\_\_\_\_, trong đó plt là bí danh của matplotlib.pyplot**

- A. plt.plot(x,y)
- B. plt.plot(y, x)
- C. plt.plot(x)
- D. Không có câu nào

ANSWER: A

**Hàm pyplot nào sau đây được sử dụng để vẽ biểu đồ histogram.**

- A. hist()
- B. histogram()
- C. histo()
- D. histochart()

ANSWER: A

**Hàm pyplot nào sau đây được sử dụng để vẽ biểu đồ hình tròn (pie chart).**

- A. pie()
- B. piechart()
- C. circle()
- D. oval()

ANSWER: A

**Hàm pyplot nào sau đây được sử dụng để vẽ biểu đồ thanh (bar graph).**

- A. bar()
- B. bargraph()
- C. barchart()
- D. oval()

ANSWER: A

**Viết một câu lệnh để hiển thị chữ “Amount” dưới dạng nhãn nằm ở trục x. (coi plt là tên bí danh của matplotlib.pyplot)**

- A. plt.xlabel(“Amount”)
- B. plt.label(“Amount”)
- C. plt.xlabel(Amount)
- D. Không có câu nào

ANSWER: A

\_\_\_\_\_ là biểu đồ cột, trong đó mỗi cột đại diện cho một phạm vi các giá trị và chiều cao của cột tương ứng với bao nhiêu giá trị trong phạm vi đó.

- A. Histograms
- B. Bar graph
- C. Line chart
- D. pie chart

ANSWER: A

**Tham số nào của hàm plot () giúp ta gán số bin của Histogram?**

- A. bins
- B. bin
- C. binvalue
- D. Không có câu nào

ANSWER: A

**Hàm nào sau đây sẽ tạo biểu đồ thanh (bar chart)?**

- A. Cả 2 câu trên
- B. bar()
- C. barh()
- D. Không có câu nào

ANSWER: A

**Đoạn mã sau sẽ hiển thị \_\_\_\_\_ dòng trong hình / biểu đồ.**  
import matplotlib.pyplot as pl

$a = [1,2,3,4,5]$   
 $b = [10, 20, 30, 40, 50]$   
 $c = [5, 10, 15, 20, 25]$   
`pl.plot(a,b)`  
`pl.plot(a,c)`  
`pl.show()`

- A. 2
- B. 1
- C. 3
- D. 4

ANSWER: A

**Một phân tích hồi quy giữa cân nặng (y tính bằng pound) và chiều cao (x tính bằng inch) cho kết quả phương trình hồi quy như sau:  $\hat{y} = 120 + 5x$ . Điều này có nghĩa là nếu chiều cao tăng thêm 1 inch, thì cân nặng dự kiến sẽ:**

- A. Tăng thêm 5 pound
- B. Tăng 1 pound
- C. Giảm 1 pound
- D. Giảm 24 pound

ANSWER: A

**Phân tích hồi quy giữa doanh số bán hàng (tính bằng \$ 1000) và quảng cáo (tính bằng \$ 100) cho kết quả phương trình hồi quy như sau:  $\hat{y} = 75 + 6x$ . Điều này có nghĩa là quảng cáo là 800 đô la, thì số tiền bán hàng dự đoán (tính bằng đô la) là:**

- A. 4875 đô la
- B. 123.000 đô la
- C. 487.500 đô la
- D. 12.300 đô la

ANSWER: A

**Phần dư (residual) hay còn gọi là sai số được định nghĩa là sự khác biệt giữa:**

- A. giá trị thực tế của y và giá trị ước tính của y
- B. giá trị thực của x và giá trị ước tính của x
- C. giá trị thực tế của y và giá trị ước tính của x
- D. giá trị thực của x và giá trị ước tính của y

ANSWER: A

**Một dấu hiệu không có mối quan hệ tuyến tính giữa hai biến sẽ là:**

- A. hệ số correlation coefficient bằng 0
- B. hệ số covariance bằng 1
- C. hệ số covariance bằng -1
- D. hệ số correlation coefficient bằng -1

ANSWER: A

**Cho các điểm dữ liệu sau (4,8), (2,5) và (1,2) và đường hồi quy  $\hat{y} = 3 + 2x$  đã được khớp với các điểm dữ liệu trên. Tổng bình phương lỗi là:**

- A. 22
- B. 10

C. 15

D. 13

ANSWER: A

**Trong phương trình hồi quy tuyến tính  $\hat{y} = a+b*x$**

A. x là biến dự đoán (predictor)

B. x là biến phản hồi (response)

C. x vừa là biến dự đoán vừa là biến phản hồi

D. y vừa là biến dự đoán vừa là biến phản hồi

ANSWER: A

**Trong phương trình hồi quy tuyến tính  $\hat{y} = a+b*x$**

A. y là biến phản hồi (response)

B. y là biến dự đoán (predictor)

C. x vừa là biến dự đoán vừa là biến phản hồi

D. y vừa là biến dự đoán vừa là biến phản hồi

ANSWER: A

**Nếu phương trình hồi quy bằng  $\hat{y} = 23,6-54,2x$ , thì 23,6 là \_\_\_\_\_ trong khi -54,2 là \_\_\_\_\_ của đường hồi quy.**

A. intercept, slope

B. slope, intercept

C. slope, regression coefficient

D. radius, intercept

ANSWER: A

**Trong phân tích hồi quy, biến được dự đoán là:**

A. biến phụ thuộc

B. biến độc lập

C. thường ký hiệu là x

D. thường ký hiệu là r

ANSWER: A

**Nếu hệ số góc của phương trình hồi quy  $y = b_0 + b_1x$  là dương thì:**

A. khi x tăng thì y tăng

B. khi x tăng thì y giảm

C. x tăng y giảm đúng hoặc x tăng y tăng đúng

D. khi x giảm thì y tăng

ANSWER: A

**Trong ứng dụng nhận dạng các chữ số viết tay từ 0 đến 9 cần xây dựng mạng nơ ron nhân tạo để huấn luyện mô hình, mạng nơ ron này cần thiết kết gồm bao nhiêu nút xuất**

A. 10

B. 2

C. 9

D. 11

ANSWER: A



**Ứng dụng nào là ứng dụng trực tiếp nhất của mạng nơ-ron?**

- A. phân loại mẫu
- B. lượng tử hóa vectơ
- C. ánh xạ mẫu
- D. các ứng dụng điều khiển

ANSWER: A

**Lan truyền ngược là một kỹ thuật học điều chỉnh trọng số trong mạng nơ-ron bằng cách lan truyền những thay đổi về trọng số.**

- A. Truyền ngược từ output tới input
- B. Truyền thẳng từ input tới output
- C. Truyền thẳng từ input tới nút ẩn
- D. Truyền ngược từ output đến nút ẩn

ANSWER: A

**Một nơ-ron nhân tạo nhận n đầu vào  $x_1, x_2, x_3, \dots, x_n$  với trọng số  $w_1, w_2, \dots, w_n$  được gắn vào các liên kết đầu vào. Tổng trọng số \_\_\_\_\_ được tính toán để chuyển tới một bộ lọc phi tuyến tính  $\Phi$  được gọi là hàm kích hoạt để giải phóng đầu ra.**

- A.  $\sum w_i * x_i$
- B.  $\sum w_i$
- C.  $\sum x_i$
- D.  $\sum w_i + \sum x_i$

ANSWER: A

**Mạng nơ-ron nhân tạo sử dụng phương pháp \_\_\_\_\_ để học**

- A. Gradient descent
- B. KNN
- C. Machine Learning
- D. Gradient ascent

ANSWER: A

**Hàm \_\_\_\_\_ nhận giá trị thực đầu vào và kết quả giá trị đầu ra là giá trị thực nằm trong khoảng từ 0 đến 1**

- A. Sigmoid
- B. Tanh
- C. ReLU
- D. Network

ANSWER: A

**Đơn vị tính toán cơ bản trong mạng nơ-ron là:**

- A. Neural
- B. Node
- C. Perceptron
- D. Unit

ANSWER: A

**Hàm kích hoạt Sigmoid  $f(t)$  được định nghĩa là:**

- A.  $\frac{1}{1+e^{-t}}$

- B.  $\frac{1}{e^t + e^{-t}}$
- C.  $t.e^{-t}$
- D.  $\frac{1}{1 + e^t}$

ANSWER: A

Hàm kích hoạt Tanh f(t) được định nghĩa là:

- A.  $\frac{e^t - e^{-t}}{e^t + e^{-t}}$
- B.  $\frac{1}{e^t + e^{-t}}$
- C.  $t.e^{-t}$
- D.  $\frac{1}{1 + e^t}$

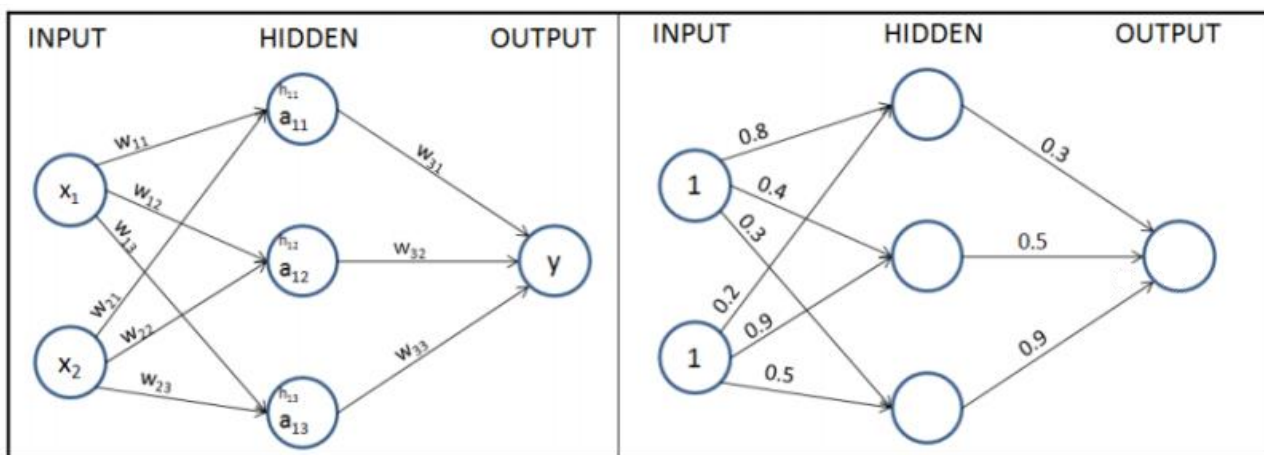
ANSWER: A

**PHẦN TỰ LUẬN (3 điểm)**

Cho mạng neural gồm 3 lớp: lớp nhập, lớp ẩn, lớp xuất với các giá trị input, trọng số được biểu diễn như hình bên dưới (các bias=0)

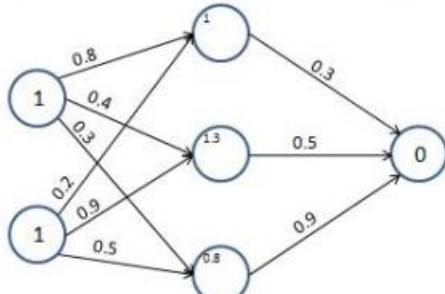
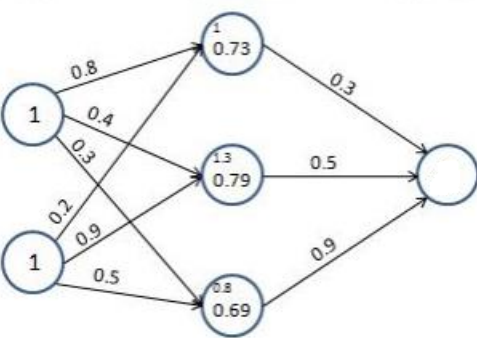
Cho hàm truyền là  $\frac{1}{1 + e^{-x}}$

Hãy tính giá trị của nút y?



**ĐÁP ÁN PHẦN TỰ LUẬN VÀ THANG ĐIỂM**

Phần câu hỏi	Nội dung đáp án	Thang điểm	Ghi chú
<b>I. Trắc nghiệm</b>		<b>7.0</b>	
Câu 1 – 50		0.14	Mỗi câu 0.14 đ
<b>II. Tự luận</b>		<b>3.0</b>	

<p><b>Câu hỏi 1</b></p>	<p> <math display="block">h_{11} = X_1 * w_{11} + X_2 * w_{21} = 1 * 0.8 + 1 * 0.2 = 1</math> <math display="block">h_{12} = X_1 * w_{12} + X_2 * w_{22} = 1 * 0.4 + 1 * 0.9 = 1.3</math> <math display="block">h_{13} = X_1 * w_{13} + X_2 * w_{23} = 1 * 0.3 + 1 * 0.5 = 0.8</math> </p> <p>             INPUT                  HIDDEN                  OUTPUT         </p>  <p> <math display="block">a_{11} = S(1.0) = \frac{1}{(1 + e^{-1})} = 0.73</math> <math display="block">a_{12} = S(-1.3) = \frac{1}{(1 + e^{-1.3})} = 0.79</math> <math display="block">a_{13} = S(-0.8) = \frac{1}{(1 + e^{-0.8})} = 0.69</math> </p> <p>             INPUT                  HIDDEN                  OUTPUT         </p>  <p> <math display="block">output\ node\ value\ (\hat{y}) = 0.73 * 0.3 + 0.79 * 0.5 + 0.69 * 0.9 = 1.235</math> </p>	<p>3.0</p>	
<p><b>Điểm tổng</b></p>		<p><b>10.0</b></p>	

TP. Hồ Chí Minh, ngày 27 tháng 11 năm 2023

Người duyệt đề

Giảng viên ra đề

(Đã duyệt)



Phùng Thế Bảo

Nguyễn Thị Mỹ Linh