

TRƯỜNG ĐẠI HỌC VĂN LANG
ĐƠN VỊ: Khoa CNTT

ĐỀ THI VÀ ĐÁP ÁN
THI KẾT THÚC HỌC PHẦN
Học kỳ 2, năm học 2023-2024

I. Thông tin chung

Tên học phần:	Nhập môn Phân tích Dữ liệu và Học sâu		
Mã học phần:	71ITDS30203	Số tín chỉ:	3
Mã nhóm lớp học phần:			
Hình thức thi: Trắc nghiệm kết hợp Tự luận	Thời gian làm bài:	60	phút
<i>Thí sinh được tham khảo tài liệu:</i>	<input type="checkbox"/> Có	<input checked="" type="checkbox"/> Không	

Gợi ý:

- SV gõ trực tiếp trên khung trả lời của hệ thống thi;

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)
CLO 1	Hiểu được nền tảng phân tích dữ liệu dựa trên các thuật toán, quy trình phát triển ứng dụng (thư viện phần mềm) dùng trong phân tích dữ liệu bằng ngôn ngữ lập trình Python	Trắc nghiệm	30%	1-20	3	PLO1/P II.2
CLO 2	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.	Tự luận	50%	Bài 1, Bài 2	5	PLO1/P II.2
CLO 3	Có kiến thức cơ bản về quy trình phân tích dữ liệu, xây dựng và khai thác các thư viện và giải thuật trong Học sâu để giải quyết một số bài toán thực tế	Trắc nghiệm + Tự luận	20%	1-20	2	PLO1/P II.2

III. Nội dung câu hỏi thi

PHẦN TRẮC NGHIỆM (tổng số câu hỏi + thang điểm từng câu hỏi)

Câu 1: Hàm nào trong NumPy được sử dụng để tạo mảng?

- A) array()
- B) create_array()
- C) new_array()
- D) make_array()

Câu 2: Phương thức nào trong Pandas được sử dụng để đọc tệp CSV vào DataFrame?

- A) import_csv()
- B) load_csv()
- C) open_csv()
- D) read_csv()

Câu 3: Làm thế nào để tìm giá trị trung bình của một mảng trong NumPy?

- A) mean()
- B) average()
- C) mean_value()
- D) mean_array()

Câu 4: Hàm nào trong Pandas được sử dụng để xử lý dữ liệu thiếu?

- A) remove_na()
- B) delete_na()
- C) omit_na()
- D) dropna()

Câu 5: Làm thế nào để tạo một DataFrame từ một từ điển trong Pandas?

- A) DataFrame.from_dict()
- B) DataFrame()
- C) DataFrame.from_dictionary()
- D) DataFrame.create_from_dict()

Câu 6: Hàm nào được sử dụng để ghép hai mảng trong NumPy?

- A) concatenate()
- B) append()

- C) merge()
- D) combine()

Câu 7: Trục mặc định cho phương thức drop trong Pandas là gì?

- A) 0 (hàng)
- B) 1 (cột)
- C) None
- D) 2 (phần tử)

Câu 8: Làm thế nào để chọn một cột cụ thể từ DataFrame trong Pandas?

- A) df.column_name
- B) df['column_name']
- C) df.select('column_name')
- D) df.column('column_name')

Câu 9: Hàm nào được sử dụng để tạo số ngẫu nhiên trong NumPy?

- A) random.rand()
- B) rand.random()
- C) generate_random()
- D) random_generate()

Câu 10: Làm thế nào để tính độ lệch chuẩn của một cột trong DataFrame trong Pandas?

- A) df['column_name'].stdev()
- B) df['column_name'].std_dev()
- C) df['column_name'].std()
- D) df['column_name'].deviation()

Câu 11: Phương thức nào trong Pandas được sử dụng để xóa các hàng trùng lặp?

- A) remove_duplicates()
- B) drop_duplicates()
- C) delete_duplicates()
- D) omit_duplicates()

Câu 12: Hàm nào trong NumPy được sử dụng để tính tích chập của hai mảng?

- A) dot()
- B) multiply()
- C) product()
- D) dot_product()

Câu 13: Làm thế nào để chuyển đổi một cột trong DataFrame sang dạng datetime trong Pandas?

- A) `pd.to_date()`
- B) `pd.to_time()`
- C) `pd.to_datetime()`
- D) `pd.datetime_convert()`

Câu 14: Phương thức nào trong Pandas có thể được sử dụng để đếm số lượng giá trị thiếu trong DataFrame?

- A) `null_count()`
- B) `missing_values()`
- C) `na_count()`
- D) `isnull().sum()`

Câu 15: Làm thế nào để lưu DataFrame vào tệp CSV trong Pandas?

- A) `to_csv()`
- B) `save_csv()`
- C) `export_csv()`
- D) `write_csv()`

Câu 16: Tình huống Xử lý dữ liệu thiếu: Bạn có một DataFrame `df` với nhiều cột chứa giá trị thiếu. Làm thế nào để bạn xử lý các giá trị thiếu này để chuẩn bị dữ liệu cho phân tích?

- A) Xóa tất cả các hàng có giá trị thiếu.
- B) Điền các giá trị thiếu bằng giá trị trung bình của cột.
- C) Xóa các cột có giá trị thiếu.
- D) Điền các giá trị thiếu bằng một giá trị cố định (ví dụ: 0).

Câu 17: Tình huống Phát hiện điểm ngoại lai (outliers): Bạn được giao một tập dữ liệu với một cột số salary. Làm thế nào để bạn phát hiện và xử lý các điểm ngoại lai trong cột này?

- A) Xóa các giá trị lớn hơn một ngưỡng cụ thể.
- B) Thay thế các điểm ngoại lai bằng giá trị trung bình.
- C) Sử dụng Z-score (outlier) để phát hiện các điểm ngoại lai và loại bỏ chúng.
- D) Bỏ qua các điểm ngoại lai.

Câu 18: Tình huống Chuẩn hóa dữ liệu: Tập dữ liệu của bạn chứa các đặc trưng với các thang đo khác nhau. Làm thế nào để bạn chuẩn hóa các đặc trưng này?

- A) Chia cho giá trị nhỏ nhất.

- B) Sử dụng Min-Max scaling.
- C) Trừ đi giá trị trung bình và chia cho độ lệch chuẩn.
- D) Để các đặc trưng như hiện tại.

Câu 19: Tình huống Chia dữ liệu: Bạn cần chia tập dữ liệu của mình thành tập huấn luyện và tập kiểm tra. Phương thức nào bạn sẽ sử dụng trong Pandas?

- A) `train_test_split()`
- B) `split_data()`
- C) `divide_data()`
- D) `split_train()`

Câu 20: Tình huống Kỹ thuật đặc trưng: Bạn có một tập dữ liệu với cột date ở dạng chuỗi. Làm thế nào để bạn tạo các đặc trưng mới như year, month, và day?

- A) Chia cột date theo các dấu phân cách.
- B) Sử dụng biểu thức chính quy để trích xuất các đặc trưng.
- C) Giữ nguyên cột date.
- D) Chuyển đổi cột date sang dạng datetime và trích xuất các đặc trưng.

PHẦN TỰ LUẬN (tổng số câu hỏi + thang điểm từng câu hỏi)

Câu hỏi 1: (2.5 điểm)

Cho input là một ảnh kích thước 4x4 và kernel (bộ lọc) kích thước 2x2. Tính toán output của phép tích chập (convolution) với stride = 1 và padding = 0.

Input	Kernel
1 2 3 4	1 0
5 6 7 8	0 -1
9 1 1 2	
3 4 5 1	

Câu hỏi 2: (2.5 điểm)

Cho tập kết quả test gồm 10 phần tử và giá trị thực tế (ground truth) tương ứng. Tính số TP, TN, FN, FP, recall, precision, và accuracy. Hãy giải thích ý nghĩa của TP, TN, FN, FP.

Predicted:	0	0	1	0	1	1	0	1	1	0
Actual (Ground truth):	0	1	1	0	1	1	1	1	0	1

Lưu ý: Sinh viên sẽ được miễn thi nếu thực hiện tiểu luận môn học và được đánh giá GIỎI

ĐÁP ÁN PHẦN TỰ LUẬN VÀ THANG ĐIỂM

Phân câu hỏi	Nội dung đáp án	Thang điểm	Ghi chú
I. Trắc nghiệm		5.0	
Câu 1 – 10	1.A 2.D 3.A 4.D 5.B 6.A 7.A 8.B 9.A 10.C	2.5	
Câu 11 – 20	11.B 12.A 13.C 14.D 15.A 16.B 17.C 18.B 19.A 20.D	2.5	
.....		5.0	
II. Tự luận		5.0	
Nội dung a.	Mỗi giá trị tính đúng đạt 0.25 điểm, -5 -5 -5 4 5 5 5 -4 0	2.0	
Nội dung b.	Đếm đúng số lượng TP, FP, TN, FN đạt 0.25 điểm; tính đúng accuracy, recall, precision đạt 0.5 điểm. <input type="checkbox"/> TP: 4 <input type="checkbox"/> TN: 2 <input type="checkbox"/> FP: 3 <input type="checkbox"/> FN: 1 Accuracy = 60%; Recall = 0.57; Precision=0.8	2.5	
		5.0	
	Điểm tổng	10.0	

TP. Hồ Chí Minh, ngày 21 tháng 07 năm 2024

Người duyệt đề

Giảng viên ra đề




Trần Ngọc Việt

Huỳnh Thái Học