CS316: Introduction to AI and Data Science
Major Exam Fall 2024

Anis Koubaa
akoubaa@psu.edu.sa
Prince Sultan University,
College of Computer and Information Sciences

Date: Monday, October 28, 2024
Duration: 60 Minutes

Instructions

- This exam allows the use of lectures notes authorized by the instructors.
- **The use of any AI tool during the exam will result in failing the course.**
- Electronic devices are not allowed except for a simple calculator.
- Read each question carefully and answer to the best of your ability.
- Write your name and student ID in the space provided below.

Student Information

Student ID:	Student Name:

Summary of Exam Questions

Category	Question Title	Time	Score	Mapped CLOs
Interview Questions	Application of Eigenvalues, Eigenvectors, and PCA in AI and Data Science	5 min	2	CLO 1, CLO 2
	Application of Dot Product, Cosine Similarity, and Normalization in NLP and Semantic Search	5 min	2	CLO 1, CLO 2
Critical-Thinking Questions	Understanding SVD in Dimensionality Reduction	5 min	2	CLO 2
	Vector Representation in Data Science	5 min	2	CLO 1
Programming Questions	Cosine Similarity Function	10 min	1	CLO 1, CLO 2
	Complete the create_face_db Function	10 min	3	CLO 1, CLO 2
	Face Recognition	10 min	3	CLO 1, CLO 2
Total for Interview Questions		10 min	4	
Total for Critical-Thinking Questions		10 min	4	
Total for Programming Questions		30 min	7	
Total Final Exam		50 min	15	

1 Interview Questions

Time: 10 minutes **Score:** 4 points **Mapped CLOs:** CLO 2

Question 1: Application of Eigenvalues, Eigenvectors, and PCA in AI and Data Science

Time: 5 minutes **Score:** 2 points **Mapped CLOs:** CLO1, CLO 2 **Word Count:** 150

PCA is an important technique in AI and Data Science.

- Explain the significance of eigenvalues and eigenvectors in the computation of the covariance matrix and their role in PCA.
- Discuss how PCA can be applied to enhance machine learning models in AI, specifically focusing on a dataset with many features, such as heath records data data or large-scale sensor data from IoT devices.

Answer:

Question 2: Application of Dot Product, Cosine Similarity, and Normalization in NLP and Semantic Search

Time: 5 minutes **Score:** 2 points **Mapped CLOs:** CLO 1, CLO 2 **Word Count:** 150

- Explain how the dot product is used in calculating cosine similarity for semantic search.
- Discuss the importance of vector normalization in the context of cosine similarity and its implications for NLP applications.
- Describe how cosine similarity enables effective semantic search in NLP applications.

Answer:

2 Critical-Thinking Questions

Time: 10 minutes **Score:** 4 points **Mapped CLOs:** CLO 1, CLO 3

For this critical thinking question, provide the correct answer along with a brief, one-sentence explanation.

Question 1: Understanding SVD in Dimensionality Reduction

Time: 5 minutes **Score:** 2 points **Mapped CLOs:** CLO 2 **Word Count:** 50

Singular Value Decomposition (SVD) is extensively used in data science, particularly for dimensionality reduction in large datasets. What is the **PRIMARY** process through which SVD achieves dimensionality reduction? **It is important to clearly explain your answer.**

A) By directly reducing the computational complexity of data processing.

B) By optimizing the storage requirements of large datasets.

C) By eliminating less significant features based purely on their variance contribution.

D) By identifying and decomposing the dataset into components of maximum variance and significance.

Answer:

Question 2: Vector Representation in Data Science

Time: 5 minutes **Score:** 2 points **Mapped CLOs:** CLO 4 **Word Count:** 50

In data science, raw data such as text or images often need to be transformed into feature vectors. What is the primary reason for representing this data in vector form? It is important to clearly explain your answer.

Answer: