

Editorial-W4A1: Analyzing Network Connectivity, PageRank Algorithm, and Support Vector Machines

Question 1

A bank's 7 regional servers were isolated due to a cyberattack. The IT head asks engineers to reconnect them with the **minimum links** to avoid delays.

Question:

How many direct connections are needed?

- A) 5
- B) 7
- C) 6
- D) 8

Correct Answer: A) 6

Explanation:

A spanning tree requires $n-1$ edges (here, n = no. of nodes = 7) to connect all nodes without loops.

Question 2

Scenario:

A startup is building a search engine and uses the PageRank algorithm. The team sets a damping factor of 0.85 to prioritize hyperlink-based navigation over random jumps.

Question:

What percentage of user interactions will involve random jumps to any page in their 4-page prototype?

- A) 5%
- B) 15%
- C) 25%
- D) 85%

Correct Answer: B) 15%

Explanation:

Random jump probability = $(1-0.85) \times 100 = 15\%$

Question 3

A news website has a "Terms of Service" page with **no outbound links**. The SEO team worries this might harm their PageRank distribution.

Question:

How will the PageRank algorithm handle this page's influence?

- A) Discard its PageRank
- B) Assign it the lowest rank

- C) Distribute its rank equally to all pages
- D) Let users decide via teleportation

Correct Answer: C) Distribute its rank equally to all pages

Explanation:

Pages with no outbound links (dangling nodes) have their PageRank redistributed evenly across all pages via the teleportation model.

Question 4 (MCQ)

An e-commerce page receives inbound links from four partner sites with PageRank scores of 0.2, 0.3, 0.4, and 0.5. Each partner site links to two pages.

Question:

What is the total PageRank contribution to the e-commerce page from these links?

- A) 0.7
- B) 1.4
- C) 2.0
- D) 0.5

Correct Answer: A) 0.7

Explanation:

Contribution = $0.2/2 + 0.3/2 + 0.4/2 + 0.5/2 = 0.7$

Question 5 (MCQ)

A blog with a PageRank of 0.9 has two outbound links. The damping factor is 0.85 to simulate user navigation behavior.

Question:

What percentage of the blog's PageRank flows through each outbound link?

- A) 42.5%
- B) 50%
- C) 85%
- D) 25%

Correct Answer: A) 42.5%

Explanation:

85% of the PageRank is distributed equally across two links: $85\% / 2 = 42.5\%$.

Question 6 (MCQ)

A telecom company is designing a fault-tolerant network to connect 6 regional data centers. The CTO insists on a topology that guarantees **no loops** and **exactly one pathway between any two centers** to avoid data packet duplication.

Question:

Which topology should the engineering team implement?

- A) Mesh Network
- B) Ring Network
- C) Tree Structure
- D) Star Network

Correct Answer: C) Tree Structure

Explanation:

A tree is an **acyclic connected graph** where any two nodes have **exactly one path** between them, satisfying the CTO's requirements.

Question 7 (MCQ)

A delivery company models cities as nodes and roads as edges. Roads have **two-way traffic** and **travel times** (e.g., 30 mins, 45 mins).

Question:

Which graph type best models this system?

- A) Directed only
- B) Undirected only
- C) Undirected and weighted
- D) Directed and unweighted

Correct Answer: C) Undirected and weighted

Explanation:

Two-way traffic requires an **undirected graph**, and travel times necessitate **weighted edges**.

Question 8

Question: Which image corresponds to lowest C value if soft margin SVM is used (Assuming Dataset is same)?

- A)
- B)
- C)
- D) Can't be determined

Correct Answer: C)

Explanation: Since margin is most in Option C image, therefore it corresponds to least C value.

Case Study for Ques. 9-15

GreenTech Inc. is a leading manufacturer of eco-friendly products. They have developed a new line of sustainable cleaning solutions and want to classify these products into 2 categories based on their environmental impact and consumer preferences. They decide to use Support Vector Machines (SVM) for this classification task.

Question 9

Question: GreenTech Inc. is using SVM to classify their eco-friendly cleaning products. What is the primary goal of SVM in this context?

- A) To minimize bias in classification
- B) To maximize the margin between different product categories
- C) To maximize variance among product features
- D) To minimize error in classification

Correct Answer: B) To maximize the margin between different product categories

Explanation: SVM aims to find the optimal hyperplane that maximizes the margin between different classes, improving generalization.

Question 10

Question: GreenTech decides to use a non-linear kernel for SVM to handle complex relationships between product features. Which kernel function is most commonly used for such tasks?

- A) Linear Kernel
- B) Polynomial Kernel
- C) Radial Basis Function (RBF) Kernel
- D) Sigmoid Kernel

Correct Answer: C) Radial Basis Function (RBF) Kernel

Explanation: The RBF kernel is widely used for mapping data to higher dimensions, making it easier to separate classes that are not linearly separable.

Question 11

Question: In the SVM model used by GreenTech, what does the C parameter control?

- A) Regularization strength
- B) Margin size
- C) Learning rate
- D) Number of support vectors

Correct Answer: A) Regularization strength

Explanation: The C parameter controls the trade-off between achieving a low error on the training set and having a large margin.

Question 12

Question: How does the RBF kernel used by GreenTech help in classifying their products?

- A) It maps data to a lower-dimensional space
- B) It computes exact Euclidean distances between data points
- C) It maps data to an infinitely high-dimensional space, creating flexible decision boundaries
- D) It is computationally less expensive than linear kernels for large datasets

Correct Answer: C) It maps data to an infinitely high-dimensional space, creating flexible decision boundaries

Explanation: The RBF kernel applies a non-linear transformation that effectively maps data into an infinitely high-dimensional space, allowing SVM to create highly flexible decision boundaries.

Question 13

Question: GreenTech is comparing the performance of `svm.SVC()` and `svm.LinearSVC()` in scikit-learn. What is the primary difference between these two?

- A) `svm.SVC()` uses an implicit linear kernel, while `svm.LinearSVC()` uses an explicit linear kernel
- B) `svm.SVC()` supports only binary classification, while `svm.LinearSVC()` supports multi-class classification by default
- C) `svm.SVC()` is designed for non-linear classification using kernels, while `svm.LinearSVC()` is designed for linear classification and optimized for large datasets
- D) `svm.SVC()` allows for only hard margin SVM, whereas `svm.LinearSVC()` allows for soft margin SVM

Correct Answer: C) `svm.SVC()` is designed for non-linear classification using kernels, while `svm.LinearSVC()` is designed for linear classification and optimized for large datasets

Explanation: `svm.SVC()` is suitable for non-linear classification using various kernels, while `svm.LinearSVC()` is optimized for linear classification and performs well with large datasets.

Question 14

Question: What is the default kernel used by `svm.SVC()` in GreenTech's implementation?

- A) Linear Kernel
- B) Polynomial Kernel
- C) Radial Basis Function (RBF) Kernel

D) Sigmoid Kernel

Correct Answer: C) Radial Basis Function (RBF) Kernel

Explanation: The default kernel in `svm.SVC()` is the Radial Basis Function (RBF) kernel, which is widely used for handling non-linear classification tasks.

Question 15

Question: What are support vectors in the context of GreenTech's SVM model?

A) Any data point used to train the model

B) Data points closest to the decision boundary, defining the margin

C) Data points furthest from the decision boundary

D) Correctly classified points with the highest margin from the decision boundary

Correct Answer: B) Data points closest to the decision boundary, defining the margin

Explanation: Support vectors are data points closest to the decision boundary and play a key role in determining the optimal margin.