Oracle AI Vector Search Practice Exam

Distance Metrics and Vector Operations - Questions

1. **Which distance metric measures the angle between two vectors?**

   ☐ Euclidean
   ☐ Manhattan
   ☒ Cosine

   Explanation: The Cosine distance metric measures the angle between two vectors by calculating the cosine of the angle between them. It is commonly used in text mining and recommendation systems where the direction of the vector is more important than its magnitude.

2. **What signifies a new era of AI capabilities in Exadata System Software 24ai?**

   ☐ Integration with open-source AI frameworks
   ☐ Automated data labeling for machine learning tasks
   ☒ Key features like AI Smart Scan and infrastructure improvements

   Explanation: Exadata System Software 24ai introduces a new era of AI capabilities through key features such as AI Smart Scan, which enhances AI processing and performance, along with significant improvements in infrastructure.

3. **You want to quickly retrieve the top-10 matches for a query vector from a dataset of billions of vectors, prioritizing speed over exact accuracy. What is the best approach?**

   ☐ Exact similarity search with a high target accuracy setting
   ☒ Approximate similarity search with a low target accuracy setting
   ☐ Exact similarity search using flat search
   ☐ Relational filtering combined with an exact search

   Explanation: Approximate similarity search (e.g., using ANN algorithms like HNSW) is designed for scenarios where speed is prioritized over exact accuracy. It efficiently retrieves approximate top matches from large datasets.

4. **What is the purpose of a vector pool in Oracle Database 23ai?**

   ☒ To store HNSW vector indexes and associated metadata
   ☐ To store all database data

☐ To store 3rd party embedding models

Explanation: The vector pool is a dedicated memory area in the System Global Area (SGA) specifically designed to store HNSW vector indexes and their metadata, optimizing vector search performance.

5. **Which SQL statement correctly adds a `VECTOR` column named `v` with 3 dimensions and `FLOAT32` format to an existing table named `my_table`?**

   ☐ `ALTER TABLE my_table MODIFY (v VECTOR(3, FLOAT32))`
   ☒ `ALTER TABLE my_table ADD (v VECTOR(3, FLOAT32))`
   ☐ `ALTER TABLE my_table ADD v VECTOR(3, FLOAT32)`
   ☐ `UPDATE my_table SET v = VECTOR(3, FLOAT32)`

   Explanation: The correct syntax for adding a new `VECTOR` column to an existing table is `ALTER TABLE table_name ADD (column_name VECTOR(dimensions, format))`. This ensures the column is properly defined with dimensions and format.

6. **What is a primary use case for GoldenGate in the context of AI?**

   ☐ To prevent data from being used by AI applications.
   ☐ To create static backups of data.
   ☒ To replicate vector changes across multiple databases.

   Explanation: GoldenGate is used to replicate and consolidate vector changes across multiple databases, enabling distributed AI processing and ensuring data consistency in AI applications.

7. **What does the FETCH APPROXIMATE clause accomplish in a similarity search?**

   ☐ Filters results by their similarity score thresholds.
   ☐ Ensures only the most accurate results are returned.
   ☒ Enables the query optimizer to use a vector index for approximate search.
   ☐ Guarantees exact matches with faster performance.

   Explanation: The `FETCH APPROXIMATE` clause allows the query optimizer to leverage a vector index, trading off some accuracy for significantly improved performance in large datasets.

8. **Which of the following is a valid dimension format for a vector data type in Oracle Database 23ai?**

   ☒ FLOAT32
   ☐ BOOLEAN

☐ STRING

Explanation: Valid dimension formats for the `VECTOR` data type include `INT8`, `FLOAT32`, `FLOAT64`, or `BINARY`. `FLOAT32` is commonly used for floating-point vector representations.

9. **Which of the following SQL queries uses Euclidean Squared Distance for a similarity search?**

   - [x] `SELECT docID FROM Vector_tab ORDER BY VECTOR_DISTANCE(embedding, query_vector, EUCLIDEAN_SQUARED) FETCH FIRST 10 ROWS ONLY;`
   - ☐ `SELECT docID FROM Vector_tab WHERE VECTOR_DISTANCE(embedding, query_vector, EUCLIDEAN) < 10;`
   - [ ] `SELECT docID FROM Vector_tab ORDER BY VECTOR_DISTANCE(embedding, query_vector, EUCLIDEAN) FETCH FIRST 10 ROWS ONLY;`
   - [ ] `SELECT docID FROM Vector_tab ORDER BY VECTOR_DISTANCE(embedding, query_vector, COSINE) FETCH FIRST 10 ROWS ONLY;`

   Explanation: The query explicitly specifies `EUCLIDEAN_SQUARED` as the distance metric, which avoids square-root calculations for faster performance while maintaining relative distance ordering.

10. **What is the key purpose of a multi-vector similarity search?**

    - ☐ Searching across multiple vector datasets simultaneously.
    - ☐ Performing searches that prioritize accuracy over speed.
    - ☐ Enhancing relational filters with vector search capabilities.
    - ☒ Finding top-k matches grouped by partitions such as documents or categories. **Explanation:** Multi-vector similarity search is designed to retrieve top-k matches within logical groupings (e.g., documents or categories), enabling context-aware results in applications like recommendation systems.

11. **What is the primary purpose of the Python code snippets provided in the context of Oracle Vector Search and OCI Generative AI?**

    - ☒ To showcase how to leverage Oracle AI Vector Search to retrieve contextually relevant information for augmenting LLM prompts
    - ☐ To explain the steps involved in setting up and configuring an Oracle Database instance for AI workloads
    - ☐ To demonstrate how to create a custom embedding model from scratch

    Explanation: The Python snippets demonstrate using Oracle AI Vector Search to retrieve relevant data for enhancing LLM prompts, enabling more accurate responses without retraining the model.

12. **Which statement accurately describes the function of the `dbms_vector_chain.util_to_embeddings()` function in Oracle Database 23ai?**

    ☐ It facilitates the compression of large text datasets for efficient storage.
    ☒ It encodes unstructured data into vector embeddings.
    ☐ It generates synthetic data to augment training datasets for AI models.

    Explanation: This function converts unstructured text into vector embeddings, enabling semantic search and AI-driven analysis.

13. **What is the primary function of AI Smart Scan in Exadata System Software 24ai?**

    ☐ To automatically optimize database queries for improved performance.
    ☐ To provide real-time monitoring and diagnostics for AI applications.
    ☒ To accelerate AI workloads by leveraging Exadata RDMA Memory (XRMEM), Exadata Smart Flash Cache, and on-storage processing.

    Explanation: AI Smart Scan offloads vector computations to storage servers, minimizing data transfer and maximizing performance for AI workloads.

14. **What happens if a `VECTOR` column contains vectors of different dimensions and an attempt is made to create a vector index on it?**

    ☐ The vectors with higher dimensions are truncated.
    ☒ An error is thrown, preventing index creation.
    ☐ The database automatically normalizes the vectors.
    ☐ The index creation succeeds, but with reduced accuracy.

    Explanation: Vector indexes require consistent dimensionality; mismatched dimensions trigger an error during creation.

15. **What is the primary purpose of using Retrieval Augmented Generation (RAG) with AI Vector Search in Oracle Database 23ai?**

    ☐ To replace traditional keyword-based search with LLM-powered semantic search.
    ☐ To store the entire LLM model within Oracle Database 23ai for faster processing.
    ☒ To provide the LLM with relevant context from Oracle Database 23ai, enabling it to generate more accurate responses.
    ☐ To train the LLM with new data from Oracle Database 23ai.

Explanation: RAG retrieves contextual data from the database to augment LLM prompts, improving response quality without model retraining.

16. **What security enhancement is introduced in Exadata System Software 24ai?**

    ☐ Integration with third-party security tools.
    ☒ SNMP Security enhancements.
    ☐ Enhanced encryption algorithms for data at rest.

    Explanation: SNMP Security improvements are a key focus in Exadata 24ai for secure network management.

17. **Which initialization parameter must be configured to enable the creation of vector indexes in Oracle Database 23ai?**

    ☐ SQL_TARGET.
    ☐ VECTOR_INDEX_REBUILD.
    ☐ QUERY_PARALLELISM.
    ☒ VECTOR_MEMORY_SIZE.

    Explanation: `VECTOR_MEMORY_SIZE` allocates memory for vector index operations.

18. **What happens when you attempt to insert a vector with an incorrect number of dimensions into a `VECTOR` column?**

    ☐ The database ignores the defined dimensions and inserts the vector as is.
    ☐ The database pads the vector with zeros to match the defined dimensions.
    ☐ The database truncates the vector to fit the defined dimensions.
    ☒ The insert operation fails, and an error message is thrown.

    Explanation: Oracle enforces strict dimensionality checks for `VECTOR` columns, rejecting mismatched data.

19. **What is a primary advantage of using approximate similarity search?**

    ☐ Guaranteed accuracy of search results.
    ☐ Support for all distance metrics equally.
    ☐ Simplified SQL syntax for vector searches.
    ☒ Faster search times in large datasets.

    Explanation: Approximate Nearest Neighbor (ANN) algorithms prioritize speed over exact accuracy, ideal for billion-scale datasets.

20. **What is a primary difference between HNSW and IVF vector indexes in Oracle Database 23ai?**

☒ HNSW supports hierarchical graphs, whereas IVF uses partition-based clustering.

☐ HNSW is used for exact searches, whereas IVF is used for approximate searches.

☐ HNSW relies on Euclidean distance only, whereas IVF supports cosine similarity only.

☐ HNSW is used for disk-based searches, whereas IVF is memory-based.

Explanation: HNSW uses graph-based hierarchical navigation, while IVF partitions data into clusters for efficient approximate search.

21. **What is created to facilitate the use of OCI Generative AI with Autonomous Database?**

☒ An AI profile for OCI Generative AI

☐ A new user account with elevated privileges

☐ A dedicated OCI compartment

Explanation: An AI profile configures the necessary settings and connections to integrate OCI Generative AI with Autonomous Database, enabling seamless interaction between the services.

22. **Which type of vector index is best suited for larger datasets that might not fit in memory?**

☐ HNSW

☐ Both HNSW and IVF are equally suitable for large datasets

☒ IVF

Explanation: IVF (Inverted File) indexes are more memory-efficient for large datasets as they partition data into clusters, unlike HNSW's memory-intensive graph structure.

23. **Which operation is NOT permitted on tables containing VECTOR columns?**

☒ `JOIN` on `VECTOR` columns

☐ `UPDATE`

☐ `DELETE`

☐ `SELECT`

Explanation: `JOIN` operations are unsupported for `VECTOR` columns due to their specialized design for similarity searches, not relational logic.

24. **In RAG, what is the role of the vector database?**

☐ To provide general knowledge to the LLM.

☐ To generate text responses directly to the user.

☒ To store private content which can be used to enhance a user's query.

Explanation: The vector database stores domain-specific or proprietary data, enabling the LLM to retrieve contextually relevant information for accurate responses.

25. **Which feature enhances RDMA over converged Ethernet (RoCE) network resilience in Exadata System Software 24ai?**

    ☐ AWR and SQL Monitor Enhancements.
    ☐ Enhanced RoCE Network Discovery.
    ☒ Improved RoCE Network Resilience.

    Explanation: Exadata 24ai introduces optimizations to RoCE network resilience, ensuring reliable high-speed data transfers for AI workloads.

26. **What is a key benefit of Select AI?**

    ☐ Increased complexity in querying data.
    ☐ A requirement to learn complex SQL queries.
    ☒ The ability to use natural language to query data.

    Explanation: Select AI simplifies database interactions by allowing users to query data using natural language instead of SQL.

27. **Which parameter adjustment could reduce indexing time for an IVF index without significantly compromising accuracy?**

    ☐ Increasing the `VECTOR_MEMORY_SIZE`.
    ☐ Using the `DOT` distance metric instead of `COSINE`.
    ☐ Setting a lower TARGET ACCURACY.
    ☒ Decreasing the number of `NEIGHBOR PARTITIONS`.

    Explanation: Reducing `NEIGHBOR PARTITIONS` speeds up IVF index creation by simplifying clustering, with a minor trade-off in granularity.

28. **What is the significance of splitting text into chunks in the process of loading data into Oracle AI Vector Search?**

    ☒ To minimize token truncation as each vector embedding model has its own maximum token limit.
    ☐ To reduce the computational burden on the embedding model.
    ☐ To facilitate parallel processing of the data during vectorization.

    Explanation: Chunking ensures text fits within the token limits of embedding models, preserving context and accuracy in vector representations.

29. **What is the advantage of using local ONNX models for embedding within the database?**

☒ Enhanced security, as no data is sent outside the database.

☐ Improved performance by leveraging cloud-based computational resources.

☐ Increased flexibility in choosing embedding models from various sources.

☐ Simplified integration with external machine learning libraries.

Explanation: Local ONNX models eliminate external data transmission, mitigating security risks while maintaining embedding quality.

30. **A developer is running a query with an HNSW index and notices that the search is returning fewer results than expected. What could be the cause?**

☐ The `VECTOR_MEMORY_SIZE` is too large.

☒ The `EFSEARCH` parameter value is too low.

☐ The query is using an incompatible distance metric.

☐ The index was created with `NEIGHBORS` set to the maximum value.

Explanation: A low `EFSEARCH` limits candidate exploration during HNSW searches, reducing result recall. Increasing it improves coverage at a minor performance cost.