Interview Questions-1

Dimensionality Reduction

(Practice Projects)

SKILLS

Easy:

Q1: What is the curse of dimensionality and why is it a problem in machine learning?

Answer: The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces that do not occur in low-dimensional settings. In machine learning, it can lead to overfitting, increased computational complexity, and the need for exponentially more data points to maintain the same level of accuracy as dimensions increase.

Q2: Explain the main difference between feature selection and feature extraction in dimensionality reduction.

Answer: Feature selection involves choosing a subset of the original features, while feature extraction creates new features by combining the original ones. PCA is an example of feature extraction, while methods like Lasso regression perform feature selection.

Q3: How does PCA work to reduce dimensionality?

Answer: PCA works by finding the principal components (orthogonal vectors) that capture the maximum variance in the data. It then projects the data onto these components, effectively reducing the number of dimensions while retaining as much of the original variance as possible.

Q4: What is the difference between PCA and LDA?

Answer: PCA is an unsupervised method that maximizes variance, while LDA is a supervised method that maximizes class separability. PCA doesn't consider class labels, whereas LDA explicitly uses them to guide the transformation.

Q5: In PCA, how do you determine the number of principal components to keep?

Answer: Common methods include:
- Using a scree plot to identify the "elbow" point
- Keeping components that explain a cumulative percentage of variance (e.g., 95%)
- Using cross-validation to select the number of components that gives the best model performance

Q6: What are eigenvectors and eigenvalues in the context of PCA?

Answer: Eigenvectors represent the directions of the principal components, while eigenvalues represent the amount of variance explained by each principal component. The eigenvector with the highest eigenvalue is the first principal component.

Q7: What is the relationship between PCA and Singular Value Decomposition (SVD)?

Answer: PCA can be implemented using SVD. When applied to the centered data matrix, the right singular vectors of SVD are equivalent to the eigenvectors found by PCA, and the singular values are the square roots of the eigenvalues.

Q8: What are some limitations of PCA?

Answer: Limitations include:
- Assumes linear relationships between features
- Can be sensitive to outliers
- May lose interpretability of features after transformation
- Might not be suitable when the variance in noise is larger than the variance in the signal

Q10: How does LDA differ from PCA in terms of optimization objective?

Answer: LDA aims to maximize the ratio of between-class variance to within-class variance, while PCA aims to maximize the overall variance in the data without considering class labels.

Medium:

Q11: Can we use PCA for feature selection?

Answer: No, PCA (Principal Component Analysis) is not a feature selection technique. PCA creates new features, called principal components, that are linear combinations of the original features. While PCA reduces dimensionality and may improve model performance, it does not result in a model that relies on a small subset of the original features. Feature selection, on the other hand, involves choosing a subset of the original features that contribute most to the model's predictive power.

Q12: How is Principal Component Analysis (PCA) used for Dimensionality Reduction?

Answer: PCA is a technique used for dimensionality reduction by projecting high-dimensional data onto a lower-dimensional space. PCA identifies the directions (principal components) that maximize the variance in the data and projects the data onto these directions. This process reduces the number of dimensions while retaining the most important information in the data, which is particularly useful when dealing with large datasets with many features.

Q13: How is the first principal component axis selected in PCA?

Answer: The first principal component axis is selected by finding the direction that maximizes the variance in the data. This axis captures the greatest amount of variability in the dataset. Mathematically, this is achieved by solving an eigenvalue problem, where the eigenvector corresponding to the largest eigenvalue is chosen as the first principal component.

Q14: What is Principal Component Analysis (PCA)?

Answer: Principal Component Analysis (PCA) is a statistical technique used to simplify a dataset by reducing its dimensions. It identifies the principal components—orthogonal vectors that capture the maximum variance in the data. By projecting the data onto these components, PCA reduces the number of dimensions while preserving as much variance as possible, making it a popular method for dimensionality reduction.

Q15: How do you perform Principal Component Analysis (PCA)?

Answer: To perform PCA, follow these steps:
- Standardize the Data: Center the data by subtracting the mean of each feature and scale if necessary.
- Compute the Covariance Matrix: Calculate the covariance matrix of the standardized data.
- Calculate the Eigenvalues and Eigenvectors: Perform eigendecomposition on the covariance matrix to obtain eigenvalues and eigenvectors.

- Sort Eigenvectors: Sort the eigenvectors by the eigenvalues in descending order.
- Select Principal Components: Choose the top k eigenvectors corresponding to the largest eigenvalues.
- Transform Data: Project the original data onto the new subspace defined by the selected principal components.
- This process reduces the dimensionality of the data while preserving as much variance as possible.

Q16: What are some advantages of using LLE over PCA?

Answer: Locally Linear Embedding (LLE) has several advantages over PCA:
- Captures Non-linear Structures: LLE can capture non-linear relationships in the data, while PCA is limited to linear correlations.
- Preserves Local Geometry: LLE preserves the local geometry of the data, making it suitable for datasets with complex, curved manifolds.
- No Need for Global Coordinates: Unlike PCA, which relies on global coordinates, LLE focuses on local neighborhood information, making it more robust for certain datasets.

Hard:

Q17: What is the difference between PCA and Random Projection approaches?

Answer:
- **PCA:** PCA identifies the directions (principal components) that maximize variance and projects the data onto these directions. It is computationally intensive but preserves the most important information in the data.
- **Random Projection:** Random Projection reduces dimensions by projecting the data onto a randomly selected set of vectors. It is computationally efficient and works well in high-dimensional spaces, but may result in some loss of information compared to PCA.

Q18: What's the difference between PCA and t-SNE?
Answer:
- **PCA:** A linear technique that preserves the global structure of the data by maximizing variance. It is suitable for linear dimensionality reduction and does not capture non-linear relationships well.
- **t-SNE:** A non-linear technique that preserves the local structure by focusing on the pairwise similarities between points. It is effective for visualizing high-dimensional data in 2D or 3D but is computationally intensive and does not provide a straightforward mapping for new data points.

Q19: Why is centering and scaling the data important before performing PCA?

Answer: Centering and scaling are important in PCA because:
- **Centering:** Ensures that the data is centered around the origin, which is crucial for accurately calculating the covariance matrix and principal components.
- **Scaling:** Prevents features with larger scales from dominating the PCA results. By scaling the data, each feature contributes equally to the analysis, ensuring a fair representation in the principal components.

Q20: Would you use PCA on large datasets, or is there a better alternative?

Answer: For large datasets, Incremental PCA (IPCA) is a better alternative. Unlike standard PCA, which requires loading all the data into memory, IPCA processes data in mini-batches, making it more memory-efficient and suitable for large datasets.

Q21: What is Sparse PCA?

Answer: Sparse PCA is a variant of PCA that introduces sparsity in the principal components. This means that some components have coefficients set to zero, making them easier to interpret. Sparse PCA is useful when you want to achieve dimensionality reduction while retaining a model that is more interpretable by focusing on a subset of features.

Q22: What is Kernel PCA?

Answer: Kernel PCA extends the PCA algorithm to non-linear datasets by using the kernel trick. It maps the data into a higher-dimensional space where linear PCA is then applied. Kernel PCA is useful for capturing non-linear relationships in the data that standard PCA cannot.

Q23: When would you use Manifold Learning techniques over PCA?

Answer: Manifold Learning techniques are preferred over PCA when dealing with datasets that have complex, non-linear structures (e.g., Swiss Roll). These techniques, like t-SNE or LLE, capture the non-linear geometry of the data, providing more meaningful low-dimensional representations for visualization or further analysis.