John Smith, Ph.D.

Department of Computer Science
University Name

Email: email@university.edu
Website: www.university.edu

July 13, 2025

John Smith, Ph.D.

Department of Computer Science
University Name

Email: email@university.edu
Website: www.university.edu

July 13, 2025

Understanding Dimensionality Reduction

What is Dimensionality Reduction?

Dimensionality reduction is the process of reducing the number of random variables under consideration, effectively simplifying a dataset while retaining its essential features.

Importance in Data Analysis

As datasets grow larger in dimensions (features), it becomes increasingly difficult to visualize, interpret, and analyze the data. Dimensionality reduction allows us to:

- **Simplify Analysis**: Focus on simplifying the analysis without losing critical information.
- **Improve Visualization**: Create meaningful visual representations (2D/3D plots).
- **Enhance Performance**: Algorithms perform more efficiently, reducing computational costs and mitigating the "curse of dimensionality."

Key Concepts

1. **Curse of Dimensionality**:
   - As the number of dimensions increases, the volume of space increases exponentially. Sparse data makes it difficult for algorithms to find patterns.
   - *Analogy*: Finding a needle in a haystack becomes more challenging as more hay (features) is added.
2. **Applications**:
   - *Image Processing*: Reducing pixel dimensions while retaining key features for better image recognition.
   - *Natural Language Processing*: Simplifying text data into meaningful representations (e.g., Word Embeddings).

Example Illustration

Data Visualization

Imagine a dataset in 4D space (e.g., properties of houses: size, price, number of rooms, location). Visualizing this data can be complex:

- **Before Reduction**: Difficult to see patterns in 4D.
- **After Reduction**: Using techniques like PCA (Principal Component Analysis), we can reduce it to a 2D plot, making trends like price vs. size easier to identify.

Summary of Key Points

- **Dimensionality Reduction** simplifies data without significant loss of information.
- It is crucial for handling **high-dimensional datasets** efficiently, promoting better analysis and visualization.
- Techniques like **PCA, t-SNE**, and **UMAP** are commonly used for this purpose in real-world applications.

By embracing dimensionality reduction, we not only make our analysis more manageable but also open the door to deeper insights in our data!

Engaging Questions to Ponder

- Have you ever faced challenges analyzing a large dataset? How did dimensionality reduction tools come into play?

- Can you think of a scenario in your everyday life where less information might deliver a clearer picture?

What is Dimensionality Reduction? - Definition

Definition

Dimensionality Reduction is a process in data analysis that reduces the number of random variables under consideration. In simpler terms, it compresses a large set of features into a smaller set while maintaining essential structures in the data.

What is Dimensionality Reduction? - Purpose

- **Simplification:** Removes unnecessary features, making data analysis and visualization easier.
- **Information Retention:** Aims to keep the most important information intact while discarding redundancy.
- **Improved Efficiency:** Reduces computation costs, enhancing the performance of machine learning models.

What is Dimensionality Reduction? - Challenges and Example

Why is This Important?

- **Curse of Dimensionality:** Increased dimensions lead to sparse data points, complicating analysis.
- **Overfitting:** More features may result in models that perform poorly on unseen data.

Example

Imagine predicting creditworthiness with features like age, income, and credit history.

- **Before:** Multiple features (age, income, etc.).
- **After (PCA):** Reduced to key principal components.

What is Dimensionality Reduction? - Key Points and Conclusion

- Aids visualization of high-dimensional data in lower-dimensional spaces (2D/3D).
- Improves training by removing noise and reducing overfitting.
- Common techniques: PCA, t-SNE, LDA.

Conclusion

Understanding dimensionality reduction helps handle large datasets effectively, maintaining data integrity while improving performance and clarity in analyses.

Why Use Dimensionality Reduction? - Introduction

Definition

Dimensionality reduction is a crucial technique in data science that simplifies datasets without sacrificing essential information.

Benefits

In this section, we will explore three key advantages:

- Improved model performance
- Reduced computational cost
- Better visualization

Why Use Dimensionality Reduction? - Improved Model Performance

- **Concept:** High-dimensional datasets can introduce noise and irrelevant features, leading to overfitting.
- **Example:** Predicting house prices using features like square footage and number of bedrooms, but including irrelevant features such as the color of the house can confuse the model.
- **Benefit:** Reducing dimensions allows models to focus on important features, improving accuracy and generalization on unseen data.

Why Use Dimensionality Reduction? - Reduced Computational Cost

- **Concept:** Processing high-dimensional data requires significant computational resources, which can be costly.
- **Example:** Training a model on a dataset with thousands of features may take days. Reducing dimensionality can significantly speed up this process.
- **Benefit:** Fewer dimensions lead to faster algorithms that require less storage space, making it feasible to work with larger datasets.

Why Use Dimensionality Reduction? - Better Visualization

- **Concept:** High-dimensional data is difficult to visualize; human interpretation is limited to 2D or 3D spaces.
- **Example:** Visualizing a dataset with 10 features is complex, but reducing dimensions to 2 allows for clear scatter plots showing clusters or patterns.
- **Benefit:** Effective visualizations facilitate understanding, identifying trends, and improving decision-making processes.

Conclusion and Code Snippet

Key Points

- Enhanced model performance by focusing on significant features and reducing noise.
- Reduced computational costs for faster processing of extensive datasets.
- Improved visualizations enable better communication of complex data insights.

Conclusion

Utilizing dimensionality reduction improves data analysis efficiency and effectiveness, driving better business decisions.

Code Snippet Example

```
from sklearn.decomposition import PCA

# Assuming X is your high-dimensional dataset
```

Common Dimensionality Reduction Techniques

Overview

Dimensionality reduction simplifies high-dimensional datasets while preserving essential characteristics. We'll explore:

- Principal Component Analysis (PCA)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Linear Discriminant Analysis (LDA)

Principal Component Analysis (PCA)

- **What It Is**: A statistical method that transforms data into a lower-dimensional space while preserving variance.
- **How It Works**: Identifies principal components based on eigenvalues and eigenvectors of the covariance matrix.

Key Points

- Reduces dimensionality by projecting onto a smaller subspace.
- Suitable for unsupervised learning and visualization.

Example

PCA can simplify thousands of pixel values in images while retaining key features like color and textures for recognition tasks.

t-SNE and LDA

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- **What It Is**: A non-linear technique for visualizing high-dimensional data in 2 or 3 dimensions.
- **How It Works**: Preserves local structure and minimizes differences in probability distributions.

Key Points

- Excellent for visualization, especially of clusters.
- More computationally intensive than PCA.

Example

Revealing clusters of similar handwritten digits in a 2D space by applying t-SNE.

Summary and Conclusion

Summary

- **PCA**: Great for unsupervised dimensionality reduction, preserves variance.
- **t-SNE**: Best for visualizing high-dimensional data; retains local relationships.
- **LDA**: Effective for supervised classification, emphasizes separability of classes.

Conclusion

Understanding these techniques is crucial for effective handling of high-dimensional data. Each method has unique advantages for specific use cases.

Principal Component Analysis (PCA)

Introduction to PCA

Principal Component Analysis (PCA) simplifies data sets by reducing their dimensionality while preserving variability. It transforms a large number of variables into a smaller set, retaining essential information. PCA is widely used in data analysis, machine learning, and image processing.

Why Use PCA?

- **Data Visualization:** Reduces high-dimensional data to 2 or 3 dimensions for easier visualization.
- **Noise Reduction:** Filters out noise, potentially reducing overfitting in predictive models.
- **Feature Extraction:** Identifies key features driving variability in data.

How PCA Works

1. **Standardize the Data:** Normalize to ensure equal contribution from each feature.
2. **Covariance Matrix:** Compute the covariance matrix to analyze relationships between variables.
3. **Eigenvalues and Eigenvectors:** Determine principal components that indicate maximum variance.
4. **Select Principal Components:** Choose top components based on eigenvalues representing variance.
5. **Transform the Data:** Project original data onto new axes defined by principal components.

Example of PCA

Before PCA

Students' scores are represented in a 4-dimensional space across subjects: Math, Science, Literature, and History.

After PCA

PCA may reduce analysis to a 2D plane where:

- 90% of variance explained by 2 principal components.
- Each point represents students' performance based on key contributing factors.

Mathematical Foundations of PCA

Covariance Matrix Calculation

$$C = \frac{1}{n-1}(X^T \times X) \tag{1}$$

Eigenvalue Equation

$$C \cdot v = \lambda \cdot v \tag{2}$$

Key Points to Emphasize

- **Dimensionality Reduction:** Simplifies data analysis by retaining significant variables.
- **Preserving Variance:** Aims to maintain original data's variance in fewer dimensions.
- **Applications:** Useful in image compression, facial recognition, and exploratory data analysis.

Conclusion

PCA is a foundational technique in data science, distilling complex datasets into actionable insights. It enables analysis bridging high-dimensional data and essential information, making it indispensable in modern data analysis.

Understanding the Mathematical Foundation of PCA

PCA (Principal Component Analysis) is a technique used for dimensionality reduction, preserving as much variance in the data as possible.

- **Dimensionality Reduction:** Reduces the number of features while retaining significant information.
- **Covariance Matrix:**
  - Describes how features co-vary.
  - Essential for identifying relationships between features.
- **Eigenvalues and Eigenvectors:**
  - Eigenvectors give the direction of the new feature space (principal components).
  - Eigenvalues provide magnitude, indicating the importance of each eigenvector.

Key Mathematical Concepts

The covariance between two variables is calculated using the formula:

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum (X_i - \bar{X})(Y_i - \bar{Y}) \tag{3}$$

where n is the number of data points, and $\bar{X}$ and $\bar{Y}$ are the means of variables X and Y. The relationship for eigenvalues and eigenvectors is given by:

$$A\mathrm{v} = \lambda\mathrm{v} \tag{4}$$

where v is an eigenvector, and λ is a corresponding eigenvalue.

How PCA Works - Part 3

Step-by-Step PCA Process

1. **Standardization:** Center data to equalize feature contributions.
2. **Covariance Matrix Calculation:** Understand feature relationships.
3. **Eigenvalues and Eigenvectors:** Reveal principal components.
4. **Sort and Select:** Choose top k eigenvectors based on largest eigenvalues.

Engagement Questions

- How might reducing dimensions impact your analysis of the dataset?
- Can you think of real-world applications where discerning essential features from complex data is crucial?

The PCA Algorithm Steps - Introduction

What is PCA?

Principal Component Analysis (PCA) is a statistical technique used to:

- Reduce the dimensionality of data

- Preserve variance

- Simplify datasets and visualize trends

- Improve machine learning model performance

Standardization

- **Purpose:** Scale data so all features contribute equally.
- **Method:**

$$Z_i = \frac{X_i - \mu}{\sigma} \tag{5}$$

  where μ is the mean and σ is the standard deviation.
- **Example:** Without standardization, weight in kilograms may dominate over height in centimeters.

1 **Covariance Matrix Computation**
  - **Purpose:** Understand relationships between features.
  - **Method:**
  $$C = \frac{1}{n-1} Z^T Z \tag{6}$$

2 **Eigenvalue and Eigenvector Calculation**
  - **Purpose:** Find principal components through eigenvalues and eigenvectors.
  - **Method: Solve:**
  $$\det(C - \lambda I) = 0 \tag{7}$$

3 **Select Principal Components**
  - **Purpose:** Choose top k eigenvectors corresponding to largest eigenvalues.
  - **Considerations:** Determine k based on explained variance ratio or domain knowledge.

The PCA Algorithm Steps - Step 5: Transformation

Transformation

- **Purpose:** Convert original data into k-dimensional space.
- **Method:**

$$Y = Z \cdot W \tag{8}$$

  where Y is the transformed data and W is the matrix of selected eigenvectors.
- **Example:** Reducing a dataset from 5 to 2 dimensions enhances interpretability.

Key Points

- PCA simplifies high-dimensional data.
- Standardization is crucial.
- Covariance reveals feature relationships guiding selections.

Choosing the Number of Principal Components - Introduction

Introduction

When performing Principal Component Analysis (PCA), selecting the optimal number of principal components (PCs) is essential for:

- Balancing dimensionality reduction with retaining sufficient information from your dataset.
- Ensuring effective analysis while avoiding information loss.

Key Considerations and Methods for Selection

Key Considerations

- **Variance Explained**: Retain PCs that explain a significant portion of the total variance to preserve essential features.
- **Overfitting vs. Underfitting**: Too few components may lead to underfitting, while too many can cause overfitting.

Methods for Selection

1. **Scree Plot**: Look for the "elbow" point in the plot of eigenvalues.
2. **Cumulative Explained Variance**: Select PCs that meet or exceed a threshold of variance (e.g., 90
3. **Cross-Validation**: Evaluate model performance for different numbers of PCs.
4. **Biological / Practical Relevance**: Use domain knowledge to select meaningful components.

Example and Key Points

Example in Context

For a dataset of flower measurements, two principal components may explain 95% of the variance. Retaining these components ensures critical information is captured.

Key Points to Emphasize

- Always plot Scree and Cumulative Variance graphs to visualize selection.
- Align selection methods with analysis goals and dataset specifics.
- Consider both statistical metrics and practical relevance for a comprehensive approach.

Visualizing PCA Results - Overview

Overview of PCA Visualization Techniques

Principal Component Analysis (PCA) is a powerful dimensionality reduction technique that simplifies high-dimensional datasets. Visualizing the results of PCA is crucial in understanding the underlying structure and relationships within the data.

- Effective methods for visualizing PCA results.
- Focus on scatter plots of principal components.

Visualizing PCA Results - Scatter Plots

1. Scatter Plots of Principal Components

Scatter plots are intuitive ways to visualize PCA results.

- By plotting the first two or three principal components, we can observe patterns and clusters.
- **Example:** A dataset of flower species with features like petal length and width.
- Plotting PC1 and PC2 reveals how species cluster together.

- Each point represents an observation (e.g., a flower).
- Axes represent variance captured by the principal components.
- Clusters indicate relationships among observations.

Visualizing PCA Results - Additional Techniques

2. Enhancing Visuals

1. Use colors or symbols to denote categories within the data.
   - **Example:** Different colors for each flower species.
2. **3. Biplots:** Combine scatter plots with original feature vectors.
   - Shows distribution and feature contributions.
   - Arrows indicate influence; longer arrows signify stronger influence.
3. **4. Explained Variance Plot:** Shows variance captured by each principal component.
   - Helps decide how many components to retain; look for "elbow" in scree plot.
4. **5. Pair Plots:** Matrix of scatter plots for each PCA pair to highlight relationships.

Visualizing PCA Results - Conclusion

Conclusion

- Visualizing PCA results is essential for understanding complex datasets.
- Techniques like scatter plots, biplots, and variance plots help identify patterns.
- Encourage students to experiment with visualizing their own PCA results to derive insights.

Additional Notes

- Emphasize interpretation: Visualization helps in making insightful conclusions.
- Encourage hands-on practice to see the impact of different visualizations.

Applications of PCA - Overview

What is PCA?

Principal Component Analysis (PCA) is a statistical technique used for dimensionality reduction. Its key goal is to simplify datasets while retaining essential characteristics.

- Transforms data into a set of linearly uncorrelated variables called principal components
- Helps uncover patterns and relationships in data
- Enhances further analysis and modeling efforts

Applications of PCA - Real-World Use Cases

1. **Image Compression**
   - Reduces dimensionality while preserving significant features.
   - Example: JPEG compression utilizes PCA for efficient image representation.

2. **Genomics and Bioinformatics**
   - Aids in visualizing genetic variation in populations with high-dimensional data.
   - Example: Gene expression studies track correlation between gene expressions and traits/diseases.

3. **Financial Market Analysis**
   - Simplifies analysis of extensive financial datasets.
   - Example: Identifies factors impacting stock market returns for effective portfolio management.

4 **Social Media Analytics**
- Reduces features from text data for theme identification.
- Example: Analyzing tweets by reducing terms to convey public sentiment efficiently.

5 **Customer Segmentation**
- Analyzes customer data to identify key segments for marketing.
- Example: E-commerce company clusters customers based on demographics and purchase histories.

Key Benefits of PCA

- Decreases data complexity, enhancing visualization and storage

- Improves model performance by reducing noise

- Versatile across various fields like finance, healthcare, and analytics

Conclusion and Reflection

Conclusion

PCA's applications cut across numerous industries, simplifying complex datasets for improved insights, model performance, and decision-making.

- Reflect on how PCA's utility can inform your analyses.
- Consider its limitations, explored in upcoming sections.

Limitations of PCA - Overview

Introduction

Principal Component Analysis (PCA) is a powerful tool for dimensionality reduction. However, it has several limitations that practitioners must be aware of:

- Linearity Assumption
- Sensitivity to Outliers
- Interpretability Challenges

Limitations of PCA - Linearity Assumption

Linearity Assumption

PCA assumes that the relationships between data points are linear. This can lead to issues when:

- As it captures only linear correlations, PCA may miss complex structures in the data.
- **Example:** Data arranged in a curved pattern (e.g., circular or parabolic) would not be well-represented by PCA.

Limitations of PCA - Sensitivity to Outliers

Sensitivity to Outliers

PCA is significantly affected by outliers. An outlier can:

- Skew the results by disproportionately influencing principal components.
- **Example:** In a dataset of student test scores, one erroneous low score can lead to misleading conclusions about overall performance.

Limitations of PCA - Interpretability

Interpretability

While PCA can reduce dimensionality, the components derived may be hard to interpret. This raises concerns for:

- Stakeholders who need to understand the implications of the results.
- **Example:** A principal component that mixes multiple test scores may not clearly represent a student's overall ability.

Key Points to Emphasize

- Always visualize data to check for non-linear patterns before applying PCA.
- Outlier detection and treatment are crucial; consider robust methods.
- While PCA simplifies data representation, it may complicate interpretability, which is vital for decision-making.

Conclusion and Next Steps

Conclusion

Understanding the limitations of PCA allows practitioners to:

- Make informed decisions on when to use PCA effectively.
- Consider exploring advanced techniques for cases where PCA may not be suitable.

Next Slide

In the next slide, we will discuss alternatives to PCA that could address its limitations.

Alternatives to PCA - Introduction

Dimensionality Reduction Techniques

While PCA (Principal Component Analysis) is a popular technique for dimensionality reduction, it has limitations such as:

- Linearity assumption
- Sensitivity to outliers

Other methods provide better performance for complex datasets. We will explore two alternatives: **t-SNE** and **UMAP**.

t-SNE (t-Distributed Stochastic Neighbor Embedding)

- **Overview:** Nonlinear dimensionality reduction for visualizing high-dimensional datasets.
- **Key Features:**
  - Maintains local structure
  - Emphasizes clusters
- **Use Cases:**
  - Visualizing high-dimensional data (e.g., NLP, genomics)
  - Exploratory data analysis

Example

Projecting a dataset of handwritten digits into two dimensions reveals clusters where '0', '1', '2', etc. form distinct clusters.

Alternatives to PCA - UMAP

UMAP (Uniform Manifold Approximation and Projection)

- **Overview:** Nonlinear method that preserves both local and global structures.
- **Key Features:**
    - Balances local and global preservation
    - Faster computation for larger datasets
- **Use Cases:**
    - High-dimensional data visualization
    - Preprocessing step for ML models

Example

UMAP can reveal segments in customer behavior datasets, helping marketers identify new customer segments.

Key Points and Conclusion

Key Points to Remember

- **t-SNE** is excellent for local structure and cluster visualization.
- **UMAP** captures both local and global structures and is faster.
- Both methods lack interpretable axes like PCA but provide insights into data structure.

Conclusion

The choice between t-SNE and UMAP depends on dataset characteristics and analysis goals. Each technique enhances understanding of complex, high-dimensional data.

Dimensionality Reduction in Practice

Key Considerations

Dimensionality reduction is a powerful technique to simplify datasets, enhance model training, and visualize complex data. However, several key considerations must be taken into account when implementing these methods in practice.

1. **Data Cleaning:**
   - Ensure the dataset is free from noise and irrelevant features.
   - Handle missing values appropriately.
   - Standardize or normalize features for consistent scaling.
   - *Example:* Normalize sensor readings to a range of [0, 1].

2. **Feature Selection:**
   - Retain only the most relevant features before applying dimensionality reduction.
   - Improves performance of reduction techniques and insights gained.

Choosing the Right Technique

- **PCA (Principal Component Analysis):**
  - Best for linear relationships and maintaining global structure.
- **t-SNE (t-distributed Stochastic Neighbor Embedding):**
  - Excels in visualizing high-dimensional data by preserving local neighborhood structures.
- **UMAP (Uniform Manifold Approximation and Projection):**
  - Balances both local and global structures, suitable for visualization while retaining overall relationships.

Validation and Computational Considerations

Validation of Results

- **Visualization:**
  - Use plots (e.g., scatter plots) to assess clusters post-reduction.
  - *Example:* Scatter plot after t-SNE reveals visible clusters.
- **Reconstruction Error:**
  - Measure how well the low-dimensional representation reconstructs the original data.

Computational Considerations

- Explore algorithms with lower computational complexity.
- Consider scalability based on expected dataset size.

Conclusion

In summary, applying dimensionality reduction involves careful:

- Preprocessing: Address noise and perform feature selection
- Choosing techniques: Align with dataset characteristics
- Validation: Use visualization and quantitative measures
- Computational considerations: Address challenges for large datasets

By thoughtfully addressing these considerations, practitioners can maximize the effectiveness of dimensionality reduction.

Tips for Implementing PCA - Overview

Principal Component Analysis (PCA)

PCA is a powerful technique used in machine learning and data analysis to reduce the dimensionality of datasets while preserving as much variability as possible. Implementing PCA effectively requires following best practices to ensure accuracy and efficiency.

Tips for Implementing PCA - Key Points

1 **Standardize Your Data:**
- Standardize your dataset (mean $= 0$, variance $= 1$) to ensure features on different scales do not disproportionately influence the analysis, as PCA is sensitive to variance.
- *Example:* Features like height and weight must be standardized.

2 **Choose the Right Number of Components:**
- Use explained variance plots (scree plots) to find the 'elbow' point for retaining components.
- *Example:* Retain components that explain significant variance (e.g., 90%).

3 **Interpret Principal Components:**
- Analyze loading vectors to understand contributions of original features and their relationships.
- *Illustration:* Strong loadings on specific features indicate underlying patterns.

4 **Use PCA for Visualization:**
- Visualize 2D or 3D reduced data to spot clusters, trends, or anomalies.
- *Example:* Scatter plots of principal components reveal class distinctions.

5 **Cross-Validate Your Results:**
- Use cross-validation to assess the impact of PCA on model performance. Compare results with and without PCA.

Engaging Questions

- Have you noticed how reducing dimensions can simplify complex datasets?
- What patterns or trends do you think will emerge after applying PCA?
- How might standardizing your data change PCA results?

Tips for Implementing PCA - Code Snippet

Here is a simple Python example using `scikit-learn` to implement PCA:

```python
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import pandas as pd

# Load dataset
data = pd.read_csv('dataset.csv')

# Standardize the data
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)

# Apply PCA
pca = PCA(n_components=2)  # Choose number of components
```

Summary of Key Points - Understanding Dimensionality Reduction

- Dimensionality reduction simplifies complex datasets by reducing the number of features while preserving essential characteristics.

- This process enhances model performance, visualization, and interpretability in machine learning.

Summary of Key Points - Key Takeaways

1. **Definition**:
   - Transform high-dimensional data into a lower-dimensional space.
2. **Importance**:
   - Enhanced Performance: Improved accuracy and faster training times.
   - Avoiding Overfitting: Reduces the likelihood of learning noise in the data.
   - Visualization: Makes patterns more apparent with lower dimensions.
3. **Common Techniques**:
   - Principal Component Analysis (PCA)
   - t-distributed Stochastic Neighbor Embedding (t-SNE)
   - Autoencoders

Summary of Key Points - Applications and Considerations

- **Applications**:
    - Image processing and analysis
    - Natural language processing
    - Genomics for biological data simplification
- **Key Considerations**:
    - Choosing the Right Technique: Select based on dataset and goals.
    - Data Preprocessing: Scale and normalize data effectively.
    - Interpreting Results: Understand transformed data for meaningful conclusions.
- **Final Thoughts**:
    - Dimensionality reduction is essential for data exploration and driving insights in machine learning.

Questions & Discussion - Part 1

Introduction to Dimensionality Reduction

Dimensionality reduction is a technique in machine learning and data analysis that simplifies complex datasets while retaining their meaningful features.

- Helps improve performance and reduce storage costs.
- Enhances data visualization.

Questions & Discussion - Part 2

Discussion Points

1 **Understanding the Need for Dimensionality Reduction**
- Challenges of high-dimensional data (the "curse of dimensionality").
- Can lead to overfitting in models.
- *Example:* Image processing with high-dimensional vectors.

2 **Common Techniques**
- **Principal Component Analysis (PCA)**: Identifies significant directions in data.
- **t-Distributed Stochastic Neighbor Embedding (t-SNE)**: Visualizes high-dimensional data in lower dimensions.

Challenges and Key Points to Emphasize

- Loss of information and limitations of dimensionality reduction.
- Importance of selecting the right method based on data type and objectives.

Questions to Spark Discussion

- What experiences have you had with dimensionality reduction?
- Can you think of scenarios where reducing dimensions could mislead interpretations?
- How might emerging techniques like Autoencoders influence practices in dimensionality reduction?