# Detailed Notes on Unsupervised Machine Learning and PCA

---

**1. Introduction to Unsupervised Machine Learning**

**Definition:**

Unsupervised Machine Learning is a type of machine learning where the model is trained on data without labeled responses. The goal is to find hidden patterns or intrinsic structures in the input data.

**Key Points:**

- **Supervised vs. Unsupervised Learning:**
  - **Supervised Learning:** Involves labeled data with input features and corresponding output labels. Examples include regression and classification.
  - **Unsupervised Learning:** Deals with unlabeled data, focusing on clustering and grouping similar data points.
- **Clustering:**
  - The primary task in unsupervised learning is clustering, where data is grouped into clusters based on similarity.
  - **Example:** Grouping customers based on purchasing behavior for targeted marketing.
- **Algorithms in Unsupervised Learning:**
  - **K-means Clustering:** Partitions data into K clusters.
  - **Hierarchical Clustering:** Builds a hierarchy of clusters.
  - **DBSCAN:** Density-based clustering for identifying outliers.
  - **Silhouette Scoring:** Used to evaluate the quality of clusters.

---

**2. Curse of Dimensionality**

**Definition:**

The curse of dimensionality refers to the challenges that arise when working with high-dimensional data, where the performance of machine learning models degrades as the number of features increases.

**Key Points:**

- **Impact of High Dimensions:**
  - As the number of features increases, the model becomes more complex and may overfit the data.
  - The model's performance decreases due to increased computational complexity and reduced generalization ability.
- **Example:**
  - Predicting house prices with 500 features. Initially, adding more features improves accuracy, but beyond a certain point, accuracy decreases due to irrelevant or redundant features.
- **Solutions to Curse of Dimensionality:**
  - **Feature Selection:** Selecting the most important features to train the model.
  - **Dimensionality Reduction:** Reducing the number of features while retaining essential information (e.g., PCA).

---

**3. Feature Selection and Extraction**

**Definition:**

- **Feature Selection:** The process of selecting a subset of relevant features for model training.
- **Feature Extraction:** The process of transforming existing features into a lower-dimensional space while retaining important information.

**Key Points:**

- **Feature Selection:**
  - **Covariance and Correlation:** Used to quantify the relationship between features.

    - **Covariance:** Measures how two variables change together.
    - **Pearson Correlation:** Ranges from -1 to 1, indicating the strength and direction of the relationship.
  - **Example:** In a housing dataset, house size is a relevant feature, while fountain size may be irrelevant.
- **Feature Extraction:**
  - **PCA (Principal Component Analysis):** A technique to reduce the number of features by creating new features (principal components) that capture the maximum variance in the data.
  - **Example:** Combining house size and number of rooms into a single feature (e.g., house size) to predict price.

---

**4. PCA Geometric Intuition**



**Definition:**

PCA is a dimensionality reduction technique that projects high-dimensional data onto a lower-dimensional space while preserving as much variance as possible.

**Key Points:**

- **Geometric Interpretation:**
  - PCA finds the best line (principal component) that captures the maximum variance in the data.
  - **Example:** In a 2D dataset, PCA finds a line where the projection of data points has the highest spread (variance).
- **Principal Components:**
  - **PC1:** The first principal component captures the maximum variance.
  - **PC2:** The second principal component captures the next highest variance and is orthogonal to PC1.
  - **Example:** In a 3D dataset, PCA finds three principal components (PC1, PC2, PC3), each capturing decreasing amounts of variance.
- **Projection:**

- Data points are projected onto the principal components to reduce dimensions while retaining maximum information.

---

**5. PCA Maths Intuition 01**

**Definition:**

PCA involves mathematical operations to find the principal components that capture the maximum variance in the data.

**Key Points:**

- **Projection and Variance:**
    - **Projection:** The process of mapping data points onto a lower-dimensional space (e.g., a line).
    - **Variance:** The spread of data points along a principal component. PCA aims to maximize this variance.
- **Cost Function:**
    - The goal of PCA is to find the unit vector (principal component) that maximizes the variance of projected data points.
    - **Equation:** The projection of a point P$P$ onto a unit vector u$u$ is given by the dot product P·u$P \cdot u$.
- **Eigenvectors and Eigenvalues:**
    - **Eigenvectors:** Represent the directions of the principal components.
    - **Eigenvalues:** Represent the magnitude of variance captured by each principal component.

---

**6. Eigen Decomposition on Covariance Matrix**

**Definition:**

Eigen decomposition is a process used in PCA to find the eigenvectors and eigenvalues of the covariance matrix, which are used to determine the principal components.

**Key Points:**

- **Covariance Matrix:**

- A matrix that captures the covariance between features.
- **Example:** For two features x*x* and y*y*, the covariance matrix is:

[Var(x)Cov(x,y)Cov(y,x)Var(y)][Var(*x*)Cov(*y*,*x*)Cov(*x*,*y*)Var(*y*)]

- **Eigenvectors and Eigenvalues:**
  - **Eigenvectors:** Represent the directions of maximum variance.
  - **Eigenvalues:** Represent the amount of variance captured by each eigenvector.
  - **Equation:** A·v=λ·v*A*·*v*=*λ*·*v*, where A*A* is the covariance matrix, v*v* is the eigenvector, and λ*λ* is the eigenvalue.
- **Steps in PCA:**
  1. **Standardize the Data:** Center the data around the mean.
  2. **Compute Covariance Matrix:** Calculate the covariance between features.
  3. **Eigen Decomposition:** Find the eigenvectors and eigenvalues of the covariance matrix.
  4. **Select Principal Components:** Choose the eigenvectors with the highest eigenvalues (PC1, PC2, etc.).
  5. **Project Data:** Project the data onto the selected principal components to reduce dimensions.

---

**Conclusion**

These notes provide a comprehensive overview of unsupervised machine learning, focusing on clustering, the curse of dimensionality, feature selection, and PCA. PCA, in particular, is a powerful technique for dimensionality reduction, leveraging geometric and mathematical intuition to capture the maximum variance in data. Eigen decomposition plays a crucial role in identifying the principal components that best represent the data in a lower-dimensional space.

# Possible interview questions related to Unsupervised Machine Learning, PCA, and dimensionality reduction

---

**1. What is Unsupervised Machine Learning?**

**Answer:**
Unsupervised learning is a type of machine learning where the model is trained on unlabeled data. The goal is to find hidden patterns or group similar data points without any predefined labels. Common tasks include clustering and dimensionality reduction.

---

**2. What is the difference between supervised and unsupervised learning?**

**Answer:**

- **Supervised Learning:** Uses labeled data (input-output pairs) to train models for tasks like classification and regression.
- **Unsupervised Learning:** Uses unlabeled data to find patterns or group data into clusters (e.g., clustering, dimensionality reduction).



---

**3. What is the curse of dimensionality?**

**Answer:**
The curse of dimensionality refers to the challenges that arise when working with high-dimensional data, such as increased computational complexity, overfitting, and reduced model performance. As the number of features grows, the data becomes sparse, making it harder for models to generalize.

---

**4. How do you handle the curse of dimensionality?**

**Answer:**

- **Feature Selection:** Select only the most important features.
- **Dimensionality Reduction:** Use techniques like PCA to reduce the number of features while retaining important information.

---

**5. What is PCA, and why is it used?**

**Answer:**
PCA (Principal Component Analysis) is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space by finding principal components (directions of maximum variance). It is used to:

- Reduce the number of features.
- Improve model performance.
- Visualize high-dimensional data in 2D or 3D.

---

**6. What are principal components in PCA?**

**Answer:**
Principal components are the directions in the data that capture the maximum variance. The first principal component (PC1) captures the most variance, the second (PC2) captures the next most, and so on. They are orthogonal (uncorrelated) to each other.

---

**7. What is the difference between feature selection and feature extraction?**

**Answer:**

- **Feature Selection:** Selecting a subset of existing features (e.g., removing irrelevant features).
- **Feature Extraction:** Creating new features from existing ones (e.g., PCA combines features into principal components).

---

**8. How does PCA work mathematically?**

**Answer:**
PCA works in the following steps:

1. **Standardize the data** (mean = 0, variance = 1).
2. **Compute the covariance matrix** to understand relationships between features.
3. **Perform eigen decomposition** on the covariance matrix to find eigenvectors (principal components) and eigenvalues (variance captured).

4. **Select the top eigenvectors** (PC1, PC2, etc.) based on eigenvalues.
5. **Project the data** onto the selected principal components to reduce dimensions.

---

**9. What are eigenvectors and eigenvalues in PCA?**

**Answer:**

- **Eigenvectors:** Directions in the data that represent principal components (PC1, PC2, etc.).
- **Eigenvalues:** Magnitudes that indicate how much variance each eigenvector captures. Higher eigenvalues mean more variance is captured.

---

**10. How do you choose the number of principal components in PCA?**

**Answer:**

- Use the **scree plot** to visualize the variance explained by each principal component.
- Choose the number of components that capture a significant amount of variance (e.g., 95% of total variance).



---

**11. What is the role of covariance in PCA?**

**Answer:**
Covariance measures how two features vary together. In PCA, the covariance matrix is used to find the relationships between features and identify the directions (principal components) that capture the most variance.

---

**12. Can PCA be used for feature selection?**

**Answer:**
No, PCA is used for **feature extraction**, not feature selection. It creates new features (principal components) by combining existing ones, rather than selecting a subset of original features.

---

**13. What is the difference between PCA and t-SNE?**

**Answer:**

- **PCA:** Linear dimensionality reduction technique that maximizes variance.
- **t-SNE:** Non-linear technique used for visualization, focusing on preserving local relationships between data points.

---

**14. What are the limitations of PCA?**

**Answer:**

- PCA assumes linear relationships between features.
- It may not work well with non-linear data.
- The principal components are harder to interpret compared to original features.

---

**15. How do you interpret principal components?**

**Answer:**
Principal components are linear combinations of original features. The weights (loadings) of each feature in a principal component indicate its contribution. Higher absolute weights mean the feature is more important for that component.



---

**16. What is the difference between PCA and LDA?**

**Answer:**

- **PCA:** Unsupervised technique that maximizes variance in the data.
- **LDA (Linear Discriminant Analysis):** Supervised technique that maximizes separation between classes.

---

**17. What is the importance of variance in PCA?**

**Answer:**
Variance represents the spread of data. PCA aims to capture the maximum variance in fewer dimensions, ensuring that the reduced data retains as much information as possible.

---

**18. Can PCA be used for classification tasks?**

**Answer:**
Yes, PCA can be used as a preprocessing step for classification tasks to reduce the number of features and improve model performance by removing noise and redundancy.

---

**19. What is the role of standardization in PCA?**

**Answer:**
Standardization (scaling features to mean = 0, variance = 1) is crucial in PCA because it ensures that features with larger scales do not dominate the principal components.

---

**20. What is the difference between PCA and SVD?**

**Answer:**

- **PCA:** Focuses on finding principal components by decomposing the covariance matrix.
- **SVD (Singular Value Decomposition):** A general matrix factorization technique that can be used to compute PCA.