**Principal Component Analysis (PCA)**

* PCA reduces the dimensionality of data by transforming it into a set of orthogonal principal components.
* The first principal component captures the maximum variance in the data.
* PCA preserves total variance; the transformation does not alter it.
* The eigenvalues of the covariance matrix represent the variance captured by each principal component.
* Principal components are uncorrelated, and their variances are given by the eigenvalues of the original data covariance matrix.
* Each principal component is a linear combination of the original variables.
* In PCA, the importance of a principal component is proportional to its eigenvalue.
* PCA can highlight underlying patterns by identifying directions (principal components) in which the data varies the most.
* PCA is effective for reducing noise in high-dimensional datasets by removing components with low eigenvalues.
* PCA is used in applications like multichannel image processing, where it can reduce redundancy across multiple spectral bands.
* PCA assumes that the principal components are ordered by the amount of variance they explain, with the first component explaining the most variance.
* PCA in multispectral imaging transforms multiple correlated spectral bands into a few independent principal components for easier interpretation.
* PCA in Face Recognition: PCA can identify "eigenfaces," which are principal components capturing variations in facial features.
* Covariance matrix properties in PCA: The covariance matrix is symmetric, with variances on the diagonal and covariances off-diagonal. Uncorrelated principal components ensure a diagonal covariance matrix in the transformed space.
* Explained variance: Eigenvalues of the covariance matrix represent the variance explained by each principal component. The fraction of total variance explained by the jthj^{th}jth principal component is , where SY​ is the covariance matrix of transformed data.
* Steps to perform PCA:
  + Center the data by subtracting the mean.
  + Compute the covariance matrix.
  + Find the eigenvalues and eigenvectors of the covariance matrix.
  + Project data onto eigenvectors to obtain principal components.
* Image compression via PCA: Multichannel images are reduced to fewer components by combining spectral bands into new images, with minimal loss of information.

### **Hypothesis Testing**

* Hypothesis tests typically involve a null hypothesis H0 and an alternative hypothesis Ha
* The null hypothesis often represents "no effect" or "no difference," while the alternative hypothesis suggests a significant effect or difference.
* A smaller p-value indicates stronger evidence against the null hypothesis.
* The p-value measures the probability of observing the test statistic, assuming the null hypothesis is true.
* The choice between a one-sided and two-sided hypothesis depends on whether the alternative hypothesis specifies a direction of difference.
* The alpha level (significance level) determines the threshold for rejecting the null hypothesis. Common values are 0.05 and 0.01.
* The test statistic in hypothesis testing is calculated differently depending on the test type (e.g., z-statistic for known variances, t-statistic for unknown variances).
* The direction of the hypothesis (one-tailed vs. two-tailed) affects the critical region of the test statistic distribution.
* P-values are typically compared against the significance level (α\alphaα) to decide whether to reject the null hypothesis.

### **Statistical Tests**

* **One-sample t-test**: Compares the sample mean to a known population mean when the population standard deviation is unknown.
* **Two-sample t-test**:
  + Compares means of two independent samples.
  + Assumes normality and independence of populations.
  + Can be applied as a paired or unpaired test depending on the relationship between samples.
* **Wilcoxon Signed-Rank Test**: A non-parametric alternative to the paired t-test used to compare medians of paired samples.
* **Wilcoxon Rank-Sum Test**: Non-parametric test for comparing medians of independent samples, also known as the Mann-Whitney U Test.
* **Chi-square Test**:
  + Test of Independence: Analyzes categorical data by examining the relationship between two variables.
  + Goodness-of-Fit Test: Tests whether the observed data distribution fits a specified distribution.
  + Assumes expected frequencies in each category are sufficiently large (commonly > 5) for reliable results.
* **Paired Samples**: Paired t-tests are used when the same group is measured under two different conditions (e.g., before and after treatment).

### **Linear Regression**

* The equation of a regression line is y = mx + b, where mmm is the slope and b is the y-intercept.
* Linear regression minimizes the sum of squared residuals (errors) to fit a line to the data.
* Residuals are the differences between observed and predicted values.
* Multicollinearity among predictors can affect the stability of regression coefficients.
* Residual plots are commonly used to check the assumptions of linear regression, such as homoscedasticity and linearity.
* A high R-squared value in linear regression indicates that a significant portion of the variance in the dependent variable is explained by the independent variables.
* Outliers can disproportionately affect regression coefficients, making robust regression techniques necessary in such cases.
* Multivariate regression extends linear regression by incorporating multiple independent variables.

### **Singular Value Decomposition (SVD)**

* SVD decomposes a matrix into three components:
  + U: Left singular vectors.
  + Σ: Diagonal matrix of singular values.
  + : Right singular vectors.
* Singular values represent the magnitude of variance captured by each component.
* Singular vectors are orthogonal and ordered by their significance in capturing variance.
* SVD is highly effective in noise reduction by reconstructing the matrix using only the most significant singular values.
* Columns of V in SVD can represent variations in handwriting styles (e.g., slant or thickness).
* SVD in image recognition: The first few singular vectors often capture the most visually significant patterns.
* Singular vectors in SVD form orthogonal bases for the row and column spaces of the matrix.
* SVD enables compression in applications like storage of large datasets by retaining only significant singular values.
* Singular images can be visualized by reshaping singular vectors into the original image dimensions in applications like digit recognition.
* Matrix reconstruction via SVD: Any matrix A can be expressed as. SVD facilitates low-rank approximations by truncating Σ to its largest singular values.

### **Vector Algebra**

* **Dot Product**:
  + Measures the similarity between two vectors based on the angle between them.
  + Calculated as
  + A dot product of zero indicates orthogonality (perpendicular vectors).
* **Cosine Similarity**:
  + Measures the cosine of the angle between two vectors, focusing on direction rather than magnitude.
  + Value ranges:
    - 1: Vectors point in the same direction.
    - 0: Vectors are orthogonal.
    - −1: Vectors point in opposite directions.
  + Invariant to the scale of the vectors, making it suitable for high-dimensional data comparisons.
* **Linear Independence**: Vectors are linearly independent if no scalar multiple of one can produce another.
* **Unit Vectors**: Represent direction independently of magnitude and have a length of 1.
* **Rotation Matrices**: Used to rotate vectors while preserving their magnitude.
* **Projection of Vectors**: The projection of one vector onto another is calculated as, where A and B are vectors.
* **Length of a Vector (L2 Norm)**: Calculated as or using the dot product
* **Parallelogram Rule**: The sum of two vectors is the diagonal of the parallelogram formed by the vectors.

### **Machine Learning**

* **Supervised Learning**: Requires labeled data to train models. Examples include classification (e.g., spam detection) and regression (e.g., housing price prediction).
* **Unsupervised Learning**: Identifies patterns in unlabeled data. Examples include clustering (e.g., grouping similar items).
* Regression models predict continuous values, while classification models predict discrete categories.
* Decision boundaries in supervised learning models like classification divide the input space into regions associated with each class.
* Feature scaling is essential for distance-based methods (e.g., k-Nearest Neighbors), as differences in feature magnitudes can disproportionately affect model performance.

### **Applications and Examples**

* **PCA in Multispectral Imaging**: Reduces multichannel images (e.g., Landsat data) to fewer dimensions while retaining most of the variance.
* **SVD in Handwriting Analysis**: Columns of V can represent "eigen-images" capturing variations in digit shapes.
* **SVD in Collaborative Filtering**: Used to decompose user-item matrices in recommender systems to predict missing ratings.
* **Chi-square in Survey Data**: Frequently used to assess whether observed responses in different demographic groups align with expected proportions.
* **Statistical Tests in Real-World Scenarios**: Comparing means of two conditions (e.g., drug effects on patient groups) or associations (e.g., gender and smoking habits) can guide practical decisions.
* **Cosine Similarity in Text Analysis**: Measures the similarity of word frequency vectors for tasks like document matching and search relevance.

### **Principal Component Analysis (PCA)**

1. PCA reduces the dimensionality of data by transforming it into a set of orthogonal principal components.
2. The first principal component captures the maximum variance in the data.
3. PCA preserves total variance; the transformation does not alter it.
4. The eigenvalues of the covariance matrix represent the variance captured by each principal component.
5. Principal components are uncorrelated, and their variances are given by the eigenvalues of the original data covariance matrix.
6. Each principal component is a linear combination of the original variables.
7. In PCA, the importance of a principal component is proportional to its eigenvalue.
8. PCA can highlight underlying patterns by identifying directions (principal components) in which the data varies the most.
9. PCA is effective for reducing noise in high-dimensional datasets by removing components with low eigenvalues.
10. PCA is used in applications like multichannel image processing, where it can reduce redundancy across multiple spectral bands.
11. PCA assumes that the principal components are ordered by the amount of variance they explain, with the first component explaining the most variance.
12. PCA in multispectral imaging transforms multiple correlated spectral bands into a few independent principal components for easier interpretation.
13. PCA in Face Recognition: PCA can identify "eigenfaces," which are principal components capturing variations in facial features.
14. Covariance matrix properties in PCA: The covariance matrix is symmetric, with variances on the diagonal and covariances off-diagonal. Uncorrelated principal components ensure a diagonal covariance matrix in the transformed space.
15. Explained variance: Eigenvalues of the covariance matrix represent the variance explained by each principal component. The fraction of total variance explained by the j-th principal component is λ\_j / tr(S\_Y), where λ\_j is the eigenvalue, and tr(S\_Y) is the trace of the covariance matrix.
16. Steps to perform PCA:

* Center the data by subtracting the mean.
* Compute the covariance matrix.
* Find the eigenvalues and eigenvectors of the covariance matrix.
* Project data onto eigenvectors to obtain principal components.

1. Image compression via PCA: Multichannel images are reduced to fewer components by combining spectral bands into new images, with minimal loss of information.
2. PCA transforms the data into a lower-dimensional space by selecting the directions (principal components) with the highest variance.
3. The orthogonal nature of PCA ensures no overlap of information between principal components, making it effective for dimensionality reduction.
4. When applied to satellite imagery (e.g., Landsat), PCA helps condense correlated spectral data into independent components for better analysis.
5. In multispectral imaging, PCA enables efficient data compression, reducing large image datasets to fewer channels with minimal loss of information.
6. PCA orders components by variance, ensuring the first few components capture the most critical data features.
7. Multispectral images benefit from PCA by condensing large datasets into fewer channels with minimal information loss.
8. PCA ensures no overlap of information between principal components by transforming data into uncorrelated dimensions.

### **Hypothesis Testing**

1. Hypothesis tests typically involve a null hypothesis (H₀) and an alternative hypothesis (Hₐ).
2. The null hypothesis often represents "no effect" or "no difference," while the alternative hypothesis suggests a significant effect or difference.
3. A smaller p-value indicates stronger evidence against the null hypothesis.
4. The p-value measures the probability of observing the test statistic, assuming the null hypothesis is true.
5. The choice between a one-sided and two-sided hypothesis depends on whether the alternative hypothesis specifies a direction of difference.
6. The alpha level (significance level) determines the threshold for rejecting the null hypothesis. Common values are 0.05 and 0.01.
7. The test statistic in hypothesis testing is calculated differently depending on the test type (e.g., z-statistic for known variances, t-statistic for unknown variances).
8. The direction of the hypothesis (one-tailed vs. two-tailed) affects the critical region of the test statistic distribution.
9. P-values are typically compared against the significance level (α) to decide whether to reject the null hypothesis.
10. Hypothesis testing evaluates the likelihood that observed patterns in sample data generalize to the population.
11. Rejecting the null hypothesis suggests the alternative hypothesis may better explain the observed data.
12. A hypothesis test evaluates whether observed differences or patterns in sample data can be generalized to the population, assuming random sampling.
13. The null hypothesis often corresponds to no effect, while the alternative suggests a meaningful difference or change in a population parameter.
14. A small p-value indicates that the observed sample statistic is unlikely under the null hypothesis, suggesting the alternative hypothesis may be more plausible.
15. Two-tailed tests are appropriate when the direction of the effect is not hypothesized, while one-tailed tests require a specific directional assumption.

### **Statistical Tests**

1. **One-Sample T-Test**:
   * Compares the sample mean to a known population mean when the population standard deviation is unknown.
   * Assumes data is drawn from a normal distribution.
   * Commonly implemented in Python using scipy.stats.ttest\_1samp.
   * The t-statistic measures how far the sample mean is from the population mean in units of standard error.
2. **Two-Sample T-Test**:
   * Compares means of two independent samples.
   * Assumes normality and independence of populations.
   * Can be applied as a paired or unpaired test depending on the relationship between samples.
   * Requires normality and equal variances for parametric applications.
   * Use scipy.stats.ttest\_ind for implementation in Python.
3. **Wilcoxon Signed-Rank Test**:
   * A non-parametric alternative to the paired t-test used to compare medians of paired samples.
4. **Wilcoxon Rank-Sum Test**:
   * Non-parametric test for comparing medians of independent samples, also known as the Mann-Whitney U Test.
   * Effective when the assumption of normality is violated.
5. **Chi-Square Test**:
   * **Test of Independence**: Analyzes categorical data by examining the relationship between two variables.
   * **Goodness-of-Fit Test**: Tests whether the observed data distribution fits a specified distribution.
   * Assumes expected frequencies in each category are sufficiently large (commonly > 5) for reliable results.
   * Always right-tailed because it measures the discrepancy between observed and expected frequencies, with larger values indicating stronger deviations from the null hypothesis.
   * Tests independence between categorical variables in contingency tables.
6. **Paired Samples**:
   * Paired t-tests analyze repeated measurements within the same sample.
   * Suitable for pre-post treatment designs or matched pair comparisons.
   * Wilcoxon signed-rank test compares medians for paired or dependent samples.

### **Linear Regression**

1. The equation of a regression line is y = mx + b, where m is the slope and b is the y-intercept.
2. Linear regression minimizes the sum of squared residuals (errors) to fit a line to the data.
3. Residuals are the differences between observed and predicted values.
4. Multicollinearity among predictors can affect the stability of regression coefficients.
5. Residual plots are commonly used to check the assumptions of linear regression, such as homoscedasticity and linearity.
6. A high R² value in linear regression indicates that a significant portion of the variance in the dependent variable is explained by the independent variables.
7. Outliers can disproportionately affect regression coefficients, making robust regression techniques necessary in such cases.
8. Multivariate regression extends linear regression by incorporating multiple independent variables.

### **Singular Value Decomposition (SVD)**

1. SVD decomposes a matrix into three components: A = UΣVᵀ.
2. U: Left singular vectors.
3. Σ: Diagonal matrix of singular values.
4. Vᵀ: Right singular vectors.
5. Singular values represent the magnitude of variance captured by each component.
6. Singular vectors are orthogonal and ordered by their significance in capturing variance.
7. SVD is highly effective in noise reduction by reconstructing the matrix using only the most significant singular values.
8. Columns of V in SVD can represent variations in handwriting styles (e.g., slant or thickness).
9. Singular vectors in SVD form orthogonal bases for the row and column spaces of the matrix.
10. SVD enables compression in applications like storage of large datasets by retaining only significant singular values.
11. The diagonal matrix Σ in SVD contains singular values ranked in decreasing order, directly correlating to the amount of variance captured by corresponding singular vectors.

### **Vector Algebra**

1. Dot Product:
   * Measures the similarity between two vectors based on the angle between them.
   * Calculated as A · B = |A||B|cos(θ).
   * A dot product of zero indicates orthogonality (perpendicular vectors).
2. Cosine Similarity:
   * Measures the cosine of the angle between two vectors, focusing on direction rather than magnitude.
   * Value ranges:
     + 1: Vectors point in the same direction.
     + 0: Vectors are orthogonal.
     + -1: Vectors point in opposite directions.
   * Invariant to the scale of the vectors, making it suitable for high-dimensional data comparisons.
   * Cosine distance is derived as 1 - Cosine Similarity, with values ranging from 0 (identical vectors) to 2 (completely dissimilar).

3. Orthogonality

* Orthogonal vectors have no projection on each other, indicated by a zero dot product.

### **Machine Learning**

1. Supervised Learning:
   * Requires labeled data to train models. Examples include classification (e.g., spam detection) and regression (e.g., housing price prediction).
2. Unsupervised Learning:
   * Identifies patterns in unlabeled data. Examples include clustering (e.g., grouping similar items).
3. SVD for Feature Selection:
   * Retaining top singular values helps identify features with predictive power while reducing overfitting risks.