Epoch Session 1 – Learning Report

Aroshish

1 Basics of Linear Algebra

1.1 Vectors and Matrices

Vectors: A vector is an ordered collection of numbers, represented as a column or row. Vectors are fundamental in linear algebra and are used to represent points in space, among other applications.
Matrices: A matrix is a rectangular array of numbers arranged in rows and columns. Operations such as addition, subtraction, and multiplication are defined for matrices and are used to solve systems of linear equations.

1.2 Matrix Operations

Addition/Subtraction: Two matrices can be added or subtracted if they have the same dimensions.
Multiplication: Matrix multiplication involves taking the dot product of rows and columns. It requires that the number of columns in the first matrix matches the number of rows in the second matrix.
Transposition: The transpose of a matrix is obtained by flipping it over its diagonal, i.e., swapping rows with columns.

1.3 Eigenvalues and Eigenvectors

Eigenvalues: For a square matrix A, an eigenvalue λ is a scalar such that there exists a non-zero vector v (eigenvector) where $Av = \lambda v$.
Eigenvectors: These are the vectors that, when multiplied by the matrix A, only get scaled by the eigenvalue λ.

2 Probability and Statistics

2.1 Basic Concepts

Probability: Measures the likelihood of an event occurring, ranging from 0 to 1.
Random Variables: A random variable is a numerical outcome of a random process. It can be discrete (countable values) or continuous (value within a range).

Probability Distributions: Describe how probabilities are distributed over the values of the random variable. (normal distribution, binomial distribution, Poisson distribution etc.).

2.2 Statistical Measures

Mean: The average of a set of values.
Median: The middle value in a sorted list of numbers.
Variance: Measures the spread of a set of values around the mean.
Standard Deviation: The square root of variance, representing the average distance of each data point from the mean.

3 Singular Value Decomposition (SVD)

3.1 Definition

SVD is a factorization method for a matrix A into three matrices: U, Σ, and V^T, where:

- U is an orthogonal matrix whose columns are left singular vectors.
- Σ is a diagonal matrix containing singular values.
- V^T is an orthogonal matrix whose rows are right singular vectors.

3.2 Applications

Dimensionality Reduction: SVD can be used to approximate the original matrix with fewer dimensions.
Image Compression: By keeping only the largest singular values, SVD can help in compressing images.

4 Dimensionality Reduction with Principal Component Analysis (PCA)

4.1 Overview

PCA is a technique used to reduce the number of variables in a dataset while preserving as much variance as possible. It works by identifying the principal components (directions of maximum variance) in the data.

4.2 Process

1. **Standardization:** Scale the data to have a mean of zero and variance of one.

2. **Covariance Matrix:** Compute the covariance matrix of the standardized data.
3. **Eigenvalue Decomposition:** Find the eigenvalues and eigenvectors of the covariance matrix.
4. **Selecting Components:** The eigenvectors with the largest eigenvalues represent the principal components.

5 Maximum Likelihood Estimation (MLE)

5.1 Definition

MLE is a method for estimating the parameters of a statistical model. It finds the parameter values that maximize the likelihood function, which measures how likely it is to observe the given data.

5.2 Process

1. **Likelihood Function:** Define the likelihood function based on the probability distribution of the data.
2. **Optimization:** Use optimization techniques to find the parameter values that maximize this function.

6 Central Limit Theorem (CLT)

6.1 Statement

The CLT states that the distribution of the sample mean of a large number of independent and identically distributed random variables approaches a normal distribution, regardless of the original distribution's shape.

6.2 Implications

Normal Approximation: CLT allows for the use of normal distribution techniques for hypothesis testing and confidence intervals, even if the data is not normally distributed.
Sample Size: The theorem becomes more accurate as the sample size increases.

7 Hypothesis Testing and p-value

7.1 Hypothesis Testing

Null Hypothesis (H0): A statement of no effect or no difference.
Alternative Hypothesis (H1): A statement that contradicts the null hypothesis.

Test Statistic: A value calculated from the sample data that is used to determine whether to reject the null hypothesis.

7.2 p-value

The p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is true.
Interpretation: A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, leading to its rejection.

8 Entropy & Information Theory

8.1 Entropy

Definition: Entropy measures the uncertainty or unpredictability of a random variable. It quantifies the average amount of information produced by a stochastic source of data.
Formula: For a discrete random variable X with probability distribution $p(x)$, entropy $H(X)$ is defined as

$$H(X) = -\sum_{x \in X} p(x) \log(p(x)).$$

8.2 Applications

Data Compression: Entropy helps in designing efficient data compression algorithms by determining the minimum number of bits needed to encode information.
Information Gain: In machine learning, entropy is used to measure the information gain in decision trees.