UNIT 2:Matrices and Related Concepts

1 Orthogonal Matrices

1.1 Definition

A matrix $Q \in \mathbb{R}^{n\times n}$ is called *orthogonal* if its transpose is equal to its inverse:

$$Q^T Q = QQ^T = I$$

where I is the identity matrix. The rows (or columns) of an orthogonal matrix are orthonormal vectors, meaning that they are:

- **Orthogonal**: The dot product of any two distinct rows (or columns) is zero.
- **Normal**: The dot product of a row (or column) with itself is 1, implying that the length of the vector is 1.

1.2 Key Properties

- Preservation of Length and Angles: If $\mathbf{x}$ is a vector, the multiplication by an orthogonal matrix preserves the Euclidean norm:

  $$\|Q\mathbf{x}\| = \|\mathbf{x}\|$$

  This property is essential for geometric transformations like rotations and reflections in $\mathbb{R}^n$.

- Inverse is the Transpose: The inverse of an orthogonal matrix is simply its transpose:

  $$Q^{-1} = Q^T$$

  This simplifies many calculations in linear algebra and numerical algorithms.

- Determinant: The determinant of an orthogonal matrix is either $+1$ or -1. If the determinant is $+1$, the transformation represented by the matrix is a pure rotation. If the determinant is -1, the transformation is a reflection.

- Eigenvalues: The eigenvalues of an orthogonal matrix lie on the unit circle in the complex plane. If Q is real, the eigenvalues are either $+1$, -1, or complex numbers of magnitude 1.

1.3 Mathematical Examples

1.3.1 Example 1: 2x2 Rotation Matrix

Consider the 2D rotation matrix:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

To verify that it is orthogonal, calculate the transpose:

$$Q^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Now, check if $Q^T Q = I$:

$$Q^T Q = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Thus, the matrix is orthogonal. This matrix represents a rotation by θ degrees in 2D.

1.3.2 Example 2: 3D Rotation Matrix

Consider the 3D rotation matrix about the z-axis:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The transpose is:

$$Q^T = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Now, verify if $Q^T Q = I$:

$$Q^T Q = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Thus, this is an orthogonal matrix, representing a rotation in 3D space.

1.3.3 Example 3: Verifying a 2x2 Orthogonal Matrix

Consider the matrix:

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

First, calculate the transpose:

$$Q^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Now check if $Q^T Q = I$:

$$Q^T Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Since $Q^T Q = I$, the matrix is orthogonal.

1.3.4 Example 4: Permutation Matrix

A permutation matrix is another example of an orthogonal matrix. Consider the matrix:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This matrix swaps the two components of a vector. The transpose of P is:

$$P^T = P$$

Multiplying P by its transpose:

$$P^T P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Thus, P is an orthogonal matrix.

1.4 Usage in Machine Learning and AI

Orthogonal matrices play a crucial role in various areas of machine learning (ML) and artificial intelligence (AI). Below are some key applications:

1.4.1 Principal Component Analysis (PCA)

In PCA, we seek an orthogonal transformation to reduce the dimensionality of the data while preserving as much variance as possible. The transformation matrix in PCA consists of the eigenvectors of the covariance matrix, which form an orthogonal matrix. These eigenvectors allow us to project the data into a new orthogonal basis where the dimensions are uncorrelated.

1.4.2 Singular Value Decomposition (SVD)

SVD is used in many machine learning algorithms, such as collaborative filtering, natural language processing, and dimensionality reduction. The SVD of a matrix A is given by:

$$A = U\Sigma V^T$$

where U and V are orthogonal matrices, and Σ is a diagonal matrix containing the singular values. The orthogonality of U and V ensures that the transformation preserves geometric properties like distances and angles between vectors.

1.4.3 QR Decomposition

QR decomposition is a factorization of a matrix A into an orthogonal matrix Q and an upper triangular matrix R:

$$A = QR$$

This decomposition is widely used in numerical algorithms, such as solving linear systems, least squares regression, and eigenvalue problems. The orthogonal matrix Q preserves numerical stability, which is crucial in machine learning algorithms involving large datasets.

1.4.4 Gradient Descent Optimization

In some advanced optimization methods, such as orthogonal gradient descent or stochastic gradient descent (SGD) on Riemannian manifolds, orthogonal matrices are used to maintain orthonormality constraints during the optimization process. This can be important in training neural networks with orthogonal weight matrices, as they help prevent issues like vanishing and exploding gradients.

1.4.5 Orthogonal Initialization in Neural Networks

In deep learning, initializing weights with orthogonal matrices has been shown to improve convergence during training. Orthogonal initialization ensures that the weight matrix does not amplify or diminish the input signals too much as they propagate through the layers, which helps in preserving the stability of the gradients.

2 Gram-Schmidt Process

The Gram-Schmidt process is a method for converting a set of linearly independent vectors into an orthonormal set.

2.1 Process

Given a set of linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, the Gram-Schmidt process generates an orthonormal set $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ as follows:

1. Set $\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$.

$$|\mathbf{v}| = \sqrt{\sum_{i=1}^{n} v_i^2}$$

2. For each $k = 2, \ldots, n$:

$$\mathbf{q}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \mathrm{proj}_{\mathbf{q}_i}(\mathbf{v}_k)$$

where the projection $\text{proj}_{\mathbf{q}_i}(\mathbf{v}_k)$ is given by:

$$\text{proj}_{\mathbf{q}_i}(\mathbf{v}_k) = \frac{\mathbf{q}_i^T \mathbf{v}_k}{\mathbf{q}_i^T \mathbf{q}_i} \mathbf{q}_i$$

3. Normalize each $\mathbf{q}_k$ after subtracting the projections.

2.2 Example

Consider two vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

1. **Step 1: Normalize $\mathbf{v}_1$:**

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

2. **Step 2: Orthogonalize $\mathbf{v}_2$ against $\mathbf{q}_1$:**

$$\text{proj}_{\mathbf{q}_1}(\mathbf{v}_2) = \frac{\mathbf{q}_1^T \mathbf{v}_2}{\mathbf{q}_1^T \mathbf{q}_1} \mathbf{q}_1 = \frac{\frac{1}{\sqrt{2}}(1+0)}{1} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Subtract the projection from $\mathbf{v}_2$:

$$\mathbf{q}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{q}_1}(\mathbf{v}_2) = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$$

3. **Step 3: Normalize $\mathbf{q}_2$:**

$$\mathbf{q}_2 = \frac{\mathbf{q}_2}{\|\mathbf{q}_2\|} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}$$

Thus, $\mathbf{q}_1$ and $\mathbf{q}_2$ form an orthonormal set.

3 Gram-Schmidt Process

Given a set of linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, the Gram-Schmidt process transforms them into an orthogonal set of vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ and then into an orthonormal set by normalizing them.

3.1 Problem Setup

Consider the following set of vectors in $\mathbb{R}^3$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

We want to use the Gram-Schmidt process to construct an orthonormal basis from these vectors.

3.2 Step 1: First Orthogonal Vector

The first orthogonal vector $\mathbf{u}_1$ is simply $\mathbf{v}_1$, as it will form the first basis vector:

$$\mathbf{u}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

3.3 Step 2: Second Orthogonal Vector

To obtain $\mathbf{u}_2$, we subtract the projection of $\mathbf{v}_2$ onto $\mathbf{u}_1$ from $\mathbf{v}_2$:

$$\mathbf{u}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_2)$$

The projection of $\mathbf{v}_2$ onto $\mathbf{u}_1$ is given by:

$$\text{proj}_{\mathbf{u}_1}(\mathbf{v}_2) = \frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1$$

First, compute the dot products:

$$\mathbf{v}_2 \cdot \mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = (1 \cdot 1) + (0 \cdot 1) + (1 \cdot 0) = 1$$

$$\mathbf{u}_1 \cdot \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = (1 \cdot 1) + (1 \cdot 1) + (0 \cdot 0) = 2$$

Now, the projection is:

$$\text{proj}_{\mathbf{u}_1}(\mathbf{v}_2) = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}$$

Now, subtract the projection from $\mathbf{v}_2$ to get $\mathbf{u}_2$:

$$\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 - \frac{1}{2} \\ 0 - \frac{1}{2} \\ 1 - 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$$

3.4 Step 3: Normalization of Vectors

To form an orthonormal basis, we need to normalize $\mathbf{u}_1$ and $\mathbf{u}_2$.

First, normalize $\mathbf{u}_1$:

$$\|\mathbf{u}_1\| = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$$

$$\hat{\mathbf{u}}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

Next, normalize $\mathbf{u}_2$:

$$\|\mathbf{u}_2\| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + 1^2} = \sqrt{\frac{1}{4} + \frac{1}{4} + 1} = \sqrt{\frac{3}{2}}$$

$$\hat{\mathbf{u}}_2 = \frac{1}{\sqrt{\frac{3}{2}}} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix} = \frac{\sqrt{2}}{\sqrt{3}} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}$$

3.5 Final Orthonormal Basis

Thus, the orthonormal basis for the subspace spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$ is:

$$\hat{\mathbf{u}}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \quad \hat{\mathbf{u}}_2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}$$