Indian Institute of Technology,
Kanpur

Important Concepts

Author:
Shubham Agrawal

February 19, 2018

Abstract

There are various concepts in probability theory and linear algebra that we keep on forgetting and they keep haunting us throughout our lifetime. I attempt to list and explain a few of them, which I myself need very much.

1 Vector Spaces

Vector Spaces are defined with respect to (set of vectors V, scalar multiplication . over F and vector addition +). A set of simple mathemetical objects form a vector space when their operations + and . follow following axioms:

$$+ : V \times V \rightarrow V$$
$$. : F \times V \rightarrow V$$

Table 1: Vector Space Properties

+	.
Associativity $u + (v + w) = (u + v) + w$	Compatibility $a(bv) = (ab)v$
Commutativity	Identity element fo scalar multiplication I.v = v
Identity Element $\exists 0 \in V \; v + 0 = v \forall v \in V$	$a(u + v) = au + av$
Inverse Element for every $v \in V, \exists v^{-1} s.t v + v^{-1} = 0$	$(a + b)v = av + bv$

2 Subspaces

Let W be a subset of vector space V (over field K). Then W is called subspace if:

- $0 \in W$
- $\forall u, v \in W \implies u + v \in W$
- $\forall c \in K$ and $w \in W \implies c.w \in W$

3 Linear Transformation

: Mapping between two vector spaces V and W (usually over same field F) such that operation is preserved. That is, applying the mapping before or after the operation should not matter.

$$f(u+v) = f(u) + f(v)$$
$$f(cu) = cf(u)$$

4 Eigen vectors and Eigen Values

Eigen Vector and Eigen values are defined with respect to Linear transformation and corresponding vector spaces. A vector is called an eigen vector wrt a particular linear trasformation if it is only scaled by a fixed value (called eigen value) on applying the linear transformation.

$$T(\mathbf{v}) = \lambda \mathbf{v}$$

There is a correspondence between n by n square matrices and linear transformations from an n-dimensional vector space to itself. For this reason, it is equivalent to define eigenvalues and eigenvectors using either the language of matrices or the language of linear transformations.

For a matrix, eigenvalues and eigenvectors can be used to decompose the matrix, for example by diagonalizing it.

5 Matrix Similarity

In linear algebra, two n-by-n matrices A and B are called similar if

$$B = P^{-1}AP$$

for some invertible n-by-n matrix P.

6 Diagonalizable Matrix

A matrix is called diagonalizable, if it's similar to a diagonal matrix.

These matrices are of interest because of their ease of handling.

7 Eigen Value decomposition

8 Singular value decomposition

9 Gram Schmidt Orthogonalization

10 Positive Definite Matrix

A symmetric (Hermitian) $n \times n$ real (complex) matrix M is said to be Positive Definite if $z^T M z > 0$ for every non-zero vector z. In case of positive semidefinite, $z^T M z \geq 0$.

10.1 Properties of Positive definite matrix

Let M be an $n \times n$ symmetric matrix.

- It has positive eigen values. Vice-versa, if a matrix has all it's eigen-values real and positive, the matrix is positive definite.
- For any real invertible matrix A, $A^T A$ **is always positive definite**
- Positive definite matrix has **positive determinant**. This means, that PD matrix is always **nonsingular**

11 Covariance Matrix

Let $\mathbf{X} = \{X_1, X_2 \cdots X_n\}$ be a random vector. Then,

$$Cov(\mathbf{X}) = [\Sigma]_{n \times n} \tag{1}$$

where $\Sigma_{ij} = cov(X_i, X_j) = E((X_i - \mu_i)(X_j - \mu_j))$
Also, Let $\mathbf{Y} = \{Y_1, Y_2 \cdots Y_m\}$ be another random vector. Then,

$$Cov(\mathbf{X},\, \mathbf{Y}) = E((\mathbf{X} - \mu_{\mathbf{X}})^T (\mathbf{Y} - \mu_{\mathbf{Y}})) = [\Sigma]_{n \times m} \tag{2}$$

11.1 Covariance

Covariance means how two things(random variables X_1 and X_2) vary with respect to each other. If on increase of X_1, X_2 increases and on decrease of X_1, X_2 decreases, we say covariance is positive. If on increase of X_1, X_2 decreases and vice versa, we say that covariance is negative. Covariance is 0 if no such relation exists. Understanding the magnitude of covariance is difficult.
Mathematically, let X_1 and X_2 are two real valued random variables. Then $\text{Cov}(X_1, X_2) = E[X_1 - \mu_1) \times (X_2 - \mu_2)]$

11.2 Properties of Covariance Matrices

- Positive semidefinite
- Symmetric
- If $(m = n)$

12 Random Variable

Random variable is a function that maps outcomes of an event to mathematically convenient form (real numbers). For example: Let our event be "*A coin is tossed 30 times*" and let random variable be **Number of heads occured while tossing 30 times**

12.1 Random Vector

Random vector are used when you want to view multiple events simultaneously. For example: You want to observe tossing of coin (X_1) and rolling of dice (X_2) simultaneously. Then, you define random vector $\mathbf{X} = \{X_1, X_2\}$. Here $X_1 : -1, 1$ and $X_2 : 1, 2, 3, 4, 5, 6$. In more machine learning pov, you have random vector as your feature set where each feature is a random variable.

13 PCA