Vectors

Oliver Zhao

1 Basic Definitions

A *vector* is an ordered finite list of numbers, usually written as a vertical array. The *elements* of the vector at the values in the array. The *size* (or *dimension* or *length*) is the number of elements contained. Two vectors are equal if the have the same size and corresponding entries. The numbers or values of the elements in a vector are called *scalars*, where the i entry in a vector a is denoted as a_i. The set of all real numbers is written as $\mathbb{R}$, while the set of all real n-vectors is denoted $\mathbb{R}^n$, where an n-vector contains n elements. We write $a \epsilon \mathbb{R}^n$ to say that a is an n-vector with real entries.

Definition 1.1 (Stacked Vectors)**.** A vector that is made by *concatenating* two or more vectors.

Definition 1.2 (Subvectors)**.** A portion of a complete vector, often notated as $a_{r:s}$ to indicate entries $a_r, \ldots, a_s$ in a size $s - r + 1$ vector.

Definition 1.3 (Notation)**.** Some documents use Greek letters to indicate numbers and low-case letters to indicate vectors. Other sources may write vectors in bold font ($\mathbf{a}$) or with an arrow ($\vec{a}$).

Definition 1.4 (Index)**.** Location within a vector. Note in computers, the index runs from $i = 0$ to $i = n - 1$ in a n-vector, while mathematical notations use $i = 1$ to $i = n$.

Definition 1.5 (Zero Vector)**.** A vector where all of its elements are zero.

Definition 1.6 (Unit Vector)**.** A vector with all elements equal to zero except one element equal to one. The ith unit vector (of size n) is the unit vector with the ith element one, denoted as e_i.

Definition 1.7 (Ones Vector)**.** A vector with all its elements equal. A n-vector that is a ones vector has the notation $\mathbf{1}_n$.

Definition 1.8 (Sparsity)**.** A vector is *sparse* if many of its entries are zero. The number of nonzero entries of an n-vector x is denoted as $\mathbf{nnz}(x)$.

2 Vector Operations

2.1 Vector Addition

Two vectors of the same size can be added together by adding the corresponding elements together to form another vector of the same size, or

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}}_{\mathbf{x+y}} \tag{1}$$

2.2 Scalar-Vector Multiplication

Scalar multiplication or *scalar-vector multiplication* is when every element within a vector is multiplied by the scalar. For example, consider a scalar β and 3-vector $\mathbf{x}$, then

$$\beta \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \beta x_1 \\ \beta x_2 \\ \beta x_3 \end{bmatrix} \tag{2}$$

Definition 2.1 (Linear Combination)**.** If $a_1, \dots, a_m$ are n-vectors and $\beta_1, \dots, \beta_m$ are scalars, the n-vector

$$\beta_1 a_1 + \cdots + \beta_m a_m \tag{3}$$

is called a *linear combination* of the vectors $a_1, \dots, a_n$. The scalars $\beta_1, \dots, \beta_m$ are called the *coefficients* of the linear combinations.

Definition 2.2 (Linear Combination of Unit Vectors)**.** Any n-vector a can be written as a linear combination of standard unit vectors in the form

$$a = a_1 e_1 + \cdots + a_n e_n \tag{4}$$

where a_i is the ith scalar element in the vector a and e_i is the i unit vector.

Definition 2.3 (Special Linear Combinations)**.** A linear combination of m vectors where the coefficients $\beta_1 = \cdots = \beta_m = 1$ is called the *sum* of vectors. A linear combination where $\beta_1 = \cdots = \beta_m = 1/m$ is called the *average* of vectors. A linear combination where the coefficients $\beta_1 + \cdots + \beta_m = 1$ is called an *affine combination*. If the coefficients in an affine combination are nonnegative, it is called a *convex combination*, a *mixture*, or a *weighted average*.

2.3 Inner Product

The *inner product* - or *dot product* - of two n-vectors is defined as the scalar

$$a^T b = a_1 b_1 + \cdots + a_n b_n. \tag{5}$$

2.4 Vector Computation Complexity

Real numbers are often stored in *floating point format* with a block of 64 bits or 8 bytes. Consequently, a n-vector requires $8n$ bytes to store. When arithmetic operations are carried out on floating point numbers, *round-off error* occurs. For simple operations, the effects are negligible. In complex models such as neural networks, it can cause significant issues.

The number of floating point operations to perform an operation is called FLOPs (or more commonly flops). The *complexity* of an operation is the number of flops required to execute it. Intuitively, performing scalar multiplication and vector addition of n-vectors require n flops, while calculating inner product of n-vectors requires $2n-1$ flops. Sparse vectors require less flops since no arithmetic operations can be reduced. For example, vector addition of the sparse vectors x and y only require $\min\{\mathbf{nnz}(x), \mathbf{nnz}(y)\}$.

3 Linear Functions

3.1 Definitions and Notations

Definition 3.1 (Function Notation)**.** The notation $f : \mathbb{R}^n \to \mathbb{R}$ means that f is a *function* that maps real n-vectors to real numbers. If n is an n-vector, then $f(x)$ denotes the *value* of the function f at x, where x is the *argument* of the function.

Definition 3.2 (Inner Product Function)**.** Suppose a is a n-vector. We can define a scalar-valued function f of n-vectors for any n-vector x, given by

$$f(x) = a^T x = a_1 x_1 + \cdots + a_n x_n \tag{6}$$

Definition 3.3 (Inner Product Representation of Linear Functions)**.** If a function is linear, then it can be expressed as the inner product of its argument with some fixed vector. Suppose f is a scalar-valued function of n-vectors and is linear for all n-vectors x, y and all scalars α, β. Then there is an n-vector a such that $f(x) = a^T x$ for all x, where $a^T x$ is called the *inner product representation* of f.

Definition 3.4 (Affine Functions)**.** A linear function plus a constant is called an *affine function*. A function $f : \mathbb{R}^n \to \mathbb{R}$ is affine if and only if it can be expressed as $f(x) = a^T x + b$ for some n-vector a and scalar b, with the latter sometimes called the *offset*.

3.2 Taylor Approximation

Scalar-valued functions of n variables, or relations between n variables and a scalar one, can often be approximated as linear or affine functions. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and z is a n-vector. The first-order *Taylor approximation* of f near (or at) point z is the function $f(x)$ at x defined as

$$\hat{f}(x) = f(z) + \frac{\partial f}{\partial x_1}(z)(x_1 - z_1) + \cdots + \frac{\partial f}{\partial x_n}(z)(x_n - z_n), \tag{7}$$

Where $\frac{\partial f}{\partial x_i}(z)$ denotes the partial derivative of f with respect to its ith argument evaluated at the n-vector z. Note that Equation (7) can be written compactly using inner product notation:

$$\hat{f}(x) = f(z) + \nabla f(z)^T (x - z). \tag{8}$$

3.3 Regression Model

A regression model is written as an affine function, where the n-vector x represents a feature vector and β is an n-vector containing the weight vector or coefficient vector, v is the offset or bias, and $\hat{y}$ is the prediction:

$$\hat{y} = x^T \beta + v. \tag{9}$$

4 Norm and Distance

4.1 Norm

Definition 4.1 (Euclidean Norm)**.** The *Euclidean norm* of an n-vector x is denoted as $\|x\|_2$ (commonly just $\|x\|$, and is defined as the square root of the sum of the squares of its elements,

$$\|x\| = \sqrt{x^T x} = \sqrt{x_1^2 + \cdots + x_n^2}. \tag{10}$$

Definition 4.2 (General Norm)**.** Any real-valued function of an n-vector that satisfies the following four properties - where x and y are equal size vectors and β is a scalar - is called a (general) norm:

- *Nonnegative homogeneity.* $\|\beta x\| = |\beta|\|x\|$
- *Triangle inequality.* $\|x + y\| \leq \|x\| + \|y\|$
- *Nonnegativity.* $\|x\| \geq 0$
- *Definiteness.* $\|x\| = 0$ only if $x = 0$

Definition 4.3 (Root-Mean-Square Value)**.** The RMS of an n-vector x is defined as

$$\mathbf{rms}(x) = \sqrt{\frac{x_1^2 + \dots x_n^2}{n}} = \frac{\|x\|}{\sqrt{n}}. \tag{11}$$

Definition 4.4 (Norm of a Sum)**.** A useful formula for the norm of the sum of two vectors x and y is

$$\|x + y\| = \sqrt{\|x\|^2 + 2x^T y + \|y\|^2}. \tag{12}$$

Definition 4.5 (Norm of Block Vectors)**.** The norm-squared of a stacked vector is the sum of the norm-squared values of its subvectors, or

$$\|(a, b, c)\| = \sqrt{\|a\|^2 + \|b\|^2 + \|c\|^2} = \|(\|a\|, \|b\|, \|c\|)\|, \tag{13}$$

Where a, b, c are vectors.

Definition 4.6 (Chebyshev Inequality)**.** Suppose x is a n-vector and that k of its entries satisfy $|x_i| \geq a$, where $a > 0$. Then k of its entries satisfy $x_i^2 \geq a^2$. Then it follows that

$$\|x\|^2 = x_1^2 + \dots + x_n^2 \geq ka^2, \tag{14}$$

Or

$$\frac{k}{n} \leq \left(\frac{\mathbf{rms}(x)}{a}\right)^2. \tag{15}$$

4.2 Distance

Definition 4.7 (Euclidean Distance)**.** The norm can be used to define the *Euclidean distance* between two vectors a and b, where

$$\mathbf{dist}(a, b) = \|a - b\|. \tag{16}$$

Definition 4.8 (Triangle Inequality)**.** Consider a triangle with vertices whose coordinates are a, b, and c. The length of any side of a triangle cannot exceed the sum of the lengths of the two other sides, or

$$\|a-c\| \leq \|a-b\| + \|b-c\|. \tag{17}$$

4.3 Standard Deviation

In this section we will use the somewhat unconventional notation $\tilde{x} = x - \mathbf{avg}(x)\mathbf{1}$ to denote the *de-meaned* vector, where each entry of x is subtracted by a constant amount such that the mean is zero. The *standard deviation* of a n-vector x is defined as the RMS value of the de-meaned vector $\tilde{x}$, or

$$\mathbf{std}(x) = \sqrt{\frac{(x_1 - \mathbf{avg}(x))^2 + \cdots + (x_n - \mathbf{avg}(x))^2}{n}} \tag{18}$$

The standard deviation can also be written using the inner product and norm as

$$\mathbf{std}(x) = \frac{\|x - (\mathbf{1}^T x/n)\mathbf{1}\|}{\sqrt{n}} \tag{19}$$

Also note the following relationship:

$$\mathbf{rms}(x)^2 = \mathbf{avg}(x)^2 + \mathbf{std}(x)^2. \tag{20}$$

4.4 Angle

Definition 4.9 (Cauchy-Schwarz Inequality)**.** For any n-vectors a and b,

$$|a^T b| \leq \|a\|\|b\|. \tag{21}$$

Definition 4.10 (Vector Angle)**.** The *angle* between two nonzero vectors a, b is defined as

$$\theta = \cos^{-1}\left(\frac{a^T b}{\|a\|\|b\|}\right). \tag{22}$$

Definition 4.11 (Norm of Sum via Angles)**.** Note that for vectors x and y we have

$$\|x+y\|^2 = \|x\|^2 + 2x^T y + \|y\| = \|x\| + 2\|x\|\|y\|\cos\theta + \|y\|. \tag{23}$$

Definition 4.12 (Correlation Coefficient)**.** The correlation between the two n-vectors a, b is

$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\|\|\tilde{b}\|} \tag{24}$$

5 Linear Independence

5.1 Definitions

Definition 5.1 (Linearly Dependant)**.** The set of n-vectors $a_1, \dots, a_k$ with $k \geq 1$ is called *linearly dependent* if

$$\beta_1 a_1 + \dots \beta_k a_k = 0 \tag{25}$$

For some $\beta_1, \dots, \beta_k$ that are not all zero. When a collection of vectors is linearly dependent, at least one of the vectors can be expressed as a linear combination of the other vectors. If $\beta_i \neq 0$ in the Equation (27), then we can express the corresponding vector a_i as

$$a_i = (-\beta_1/\beta_i)a_1 + \dots + (-\beta_{i-1}/\beta_i)a_{i-1} + (-\beta_{i+1}\beta_i)a_{i+1} + \dots + (-\beta_k\beta_i)a_k. \tag{26}$$

Definition 5.2 (Linearly Independent)**.** The set of n-vectors $a_1, \dots, a_k$ with $k \geq 1$ is called *linearly independent* if

$$\beta_1 a_1 + \dots \beta_k a_k = 0 \tag{27}$$

Is only true if $\beta_1 = \dots = \beta_k = 0$.

Definition 5.3 (Linear Combination of Linearly Independent Vectors)**.** Suppose a vector x is a linear combination of $a_1, \dots, a_k$, or

$$x = \beta_1 a_1 + \dots + \beta_k a_k. \tag{28}$$

When the vectors $a_1, \dots, a_k$ are linearly independent, the coefficients that form x are unique. In fact, the set of these unique coefficients are a unique set for any vector x that is a linear combination of the set of linearly independent vectors $a_1, \dots, a_k$.

Definition 5.4 (Supersets and Subsets)**.** If a collection of vectors is linearly dependent, any superset of it is linearly dependent. If a collection of vectors is linearly independent, any subset of it is also linearly independent.

5.2 Basis

Definition 5.5 (Independence-Dimension Inequality)**.** If the n-vectors $a_1, \dots, a_k$ are linearly independent, then $k \leq n$.

Definition 5.6 (Basis)**.** A collection of n linearly independent n-vectors is called a *basis*. Any n-vector b can be written as a linear combination of the vectors within a basis.

Definition 5.7 (Expansion in a Basis)**.** When we express an n-vector b as a linear combination of a basis $a_1, \ldots, a_n$, we refer to

$$b = \alpha_1 a_1 + \cdots + \alpha_n a_n, \tag{29}$$

As the *expansion of* b *in the* $a_1, \ldots, a_n$ *basis.* The numbers $\alpha_1, \ldots, \alpha_n$ are called the *coefficients* of the expansion of b in the basis $a_1, \ldots, a_n$.

5.3 Orthonormal Vectors

A collection of vectors $a_1, \ldots, a_k$ is *orthogonal* or *mutually orthogonal* if $a_i \perp a_j$ for any i, j with $i \neq j$, $i, j = 1, \ldots, k$. A collection of vectors $a_1, \ldots, a_k$ is *orthonormal* if it is orthogonal and $\|a_i\| = 1$ for $i = 1, \ldots, k$.

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{30}$$

Note that orthonormal vectors are linearly independent. Meanwhile, if the n-vectors $a_1, \ldots, a_n$ are orthonormal and linearly independent, they also form an *orthonormal basis.* If $a_1, \ldots, a_n$ are an orthonormal basis, then we have for any n-vector x, the identity

$$x = (a_1^T x)a_1 + \cdots + (a_n^T x)a_n. \tag{31}$$

5.4 Gram-Schmidt Algorithm

If the set of n-vectors $a_1, \ldots, a_k$ are linearly independent, the Gram-Schmidt algorithm produces an orthonormal collection of vectors $q_1, \ldots, q_k$ with the following properties:

For each $i = 1, \ldots, k$*,* a_i *is a linear combination of* $q_1, \ldots, q_i$ *and* q_i *is a linear combination of* $a_1, \ldots, a_i$.

If the vectors $a_1, \ldots, a_{j-1}$ are linearly independent, but $a_1, \ldots, a_j$ are linearly dependent, the algorithm detects this and terminates. In short, the Gram-Schmidt algorithm finds the first vector a_j that is a linear combination of the previous vectors $a_1, \ldots, a_{j-1}$.

Gram-Schmidt Algorithm

Given n-vectors $a_1, \ldots, a_k$:

for $i = 1, \ldots, k$

1. *Orthogonalization.*
$$\tilde{q} = a_i - (q_1^T a_i)q_1 - \cdots - (q_{i-1}^T a_i)q_{i-1} \tag{32}$$
2. *Test for linear dependence.*
$$\text{if } \tilde{q}_i = 0 \text{ quit} \tag{33}$$
3. *Normalization.*
$$q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|} \tag{34}$$

If the algorithm successfully runs to completion, then the original collection of vectors are linearly independent.

Determining if a vector is a linear combination of linearly independent vectors. Suppose the vectors $a_1, \ldots, a_k$ are linearly independent and we want to determine if another vector b is a linear combination of them. We can run the algorithm to the list of $k+1$ vectors $a_1, \ldots, a_k, b$. These vectors are linearly dependent if b is a linear combination of $a_1, ;a_k$ and the algorithm will terminate in the $(k+1)$st step. Otherwise, the algorithm will run to completion.

Checking if a collection of vectors is a basis. To check if the n-vectors $a_1, \ldots, a_n$ are a basis, we can run the Gram-Schmidt algorithm. If the algorithm runs to completion, the collection of vectors is a basis.