Vectors

Vectors and the scalar multiplication and vector addition operations:

$$\mathbf{x}_{n\times 1} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \quad 2\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} + 3\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 2x_1 + 3y_1 \\ 2x_2 + 3y_2 \\ \dots \\ 2x_n + 3y_n \end{pmatrix}$$

I'll use the two terms "vector" and "point" interchangeable: any point $\in \mathbb{R}^n$ corresponds to a vector starting from the origin and ending at that point.

- The inner (dot or cross) product of two vectors is defined to be

$$\mathbf{u}^t\mathbf{v} = \sum_i u_i v_i = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cos(\theta),$$

  where $\|\mathbf{u}\|$ denotes the norm of a vector

$$\|\mathbf{u}\| = \sqrt{\mathbf{u}^t\mathbf{u}} = \sqrt{\sum_i u_i^2},$$

  and θ is the angle between the two vectors.

- A unit vector is a vector whose norm is 1.

- When two vectors are orthogonal, we mean $\cos(\theta) = 0$, therefore $\mathbf{u}^t\mathbf{v} = 0$, denoted by $\mathbf{u} \perp \mathbf{v}$.

- The Euclidean distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ is $\|\mathbf{u} - \mathbf{v}\|$.

Linear Combinations & Spans

- A linear combination of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_p$ is

$$b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_p\mathbf{x}_p, \quad b_1, \ldots, b_p \in \mathbb{R}.$$

- $\mathsf{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_p)$ denotes the set of all linear combinations of the p vectors, $\mathbf{x}_1, \ldots, \mathbf{x}_p$.

- Concatenate the p vectors into a matrix $\mathbf{X}_{n\times p} = (\mathbf{x}_1 \cdots \mathbf{x}_p)$. The column space of $\mathbf{X}$ is

$$C(\mathbf{X}) = \mathsf{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_p).$$

  Any vector in $C(\mathbf{X})$ can be written as $\mathbf{X}_{n\times p}\mathbf{b}_{p\times 1}$, where $\mathbf{b} = (b_1, \ldots, b_p)^t$.

Linear Subspace

- Let $\mathcal{M}$ be a collection of vectors from $\mathbb{R}^n$. $\mathcal{M}$ is a vector space if $\mathcal{M}$ is closed under linear combinations, that is,

$$\text{if } \mathbf{u}, \mathbf{v} \in \mathcal{M}, \text{ then } a\mathbf{u} + b\mathbf{v} \in \mathcal{M},$$

  where $a, b \in \mathbb{R}$.

- You can image a linear subspace as a bag of vectors, and for any two vectors in of that bag, say $\mathbf{u}$ and $\mathbf{v}$ (the two vectors could be the same, i.e., you are allowed to create copies of vectors in that bag), their linear combination, say $\mathbf{u} - 2\mathbf{v}$, should also be in that bag.

- Apparently, we have $\mathbf{u} - \mathbf{u} = \mathbf{0}$, so $\mathbf{0}$ is in any linear subspace. (i.e., any linear subspace should pass the origin).

Examples of vector spaces

- $\mathbf{0}$ (the smallest subspace)
- $\mathbb{R}^n$ (the largest one)
- Collection of all points on a line passing through the origin
- A span of $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ is a linear subspace
- The column space of a matrix $\mathbf{X}_{n \times p}$, $C(\mathbf{X})$. Apparently, if we change the order of the columns of $\mathbf{X}$, the column space $C(\mathbf{X})$ remains the same.
- $\mathcal{M}$ and $\mathcal{N}$ are vector spaces, so are $\mathcal{M} \cap \mathcal{N}$ and $\mathcal{M} + \mathcal{N} = \{\mathbf{u} + \mathbf{v}, \mathbf{u} \in \mathcal{M}, \mathbf{v} \in \mathcal{N}\}$. But $\mathcal{M} \cup \mathcal{N}$ may not.

Replacement Rule

- Given p vectors: $\mathbf{x}_1, \ldots, \mathbf{x}_p$, define

$$\mathbf{z}_1 = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_p\mathbf{x}_p.$$

If $a_1 \neq 0$, then the two follow subspaces are the same

$$\mathsf{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p) = \mathsf{Span}(\mathbf{z}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p).$$

- Similarly, let $\tilde{\mathbf{X}}$ be a new matrix which is the same as $\mathbf{X}$ except that we replace the jth column by a linear combination,

$$\tilde{\mathbf{X}}[,j] = a_j\mathbf{X}[,j] + \sum_{i \neq j} a_i\mathbf{X}[,i].$$

If $a_j \neq 0$, then $C(\tilde{\mathbf{X}}) = C(\mathbf{X})$.

Orthogonality

- Vector $\perp$ Vector: $\mathbf{u} \perp \mathbf{v}$ if $\mathbf{u}^t\mathbf{v} = 0$.

- Vector $\perp$ Subspace: $\mathbf{u} \perp \mathcal{M}$, if $\mathbf{u}$ is orthogonal to any vector from $\mathcal{M}$. For example, if $\mathbf{u}$ is orthogonal to each column of a matrix $\mathbf{X}$, then we have $\mathbf{u} \perp C(\mathbf{X})$.

- Subspace $\perp$ Subspace: Similarly we can define $\mathcal{M} \perp \mathcal{N}$, if any vector from $\mathcal{M}$ is orthogonal to any vector from $\mathcal{N}$.

- Define the orthogonal complement of subspace V as

$$\mathcal{M}^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \perp \mathcal{M}\}.$$

It is easy to show that $\mathcal{M}^\perp$ is a subspace. Apparently $\mathcal{M} \perp \mathcal{M}^\perp$.

Linear Independence

- A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is said to be linear independent, if

$$c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m = 0 \text{ iff } c_1 = \cdots = c_m = 0.$$

  Otherwise they are linear dependent.

- In other words, if a set of vectors are linear independent, then **no one** can be expressed as a linear combination of the others; if they are linear dependent, then there **exists one** vector, say $\mathbf{v}_2$, which can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_3, \ldots, \mathbf{v}_m$.

Linear Independence and Bases

A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a basis for a subspace $\mathcal{M}$, if

1. Span$(\mathbf{u}_1, \ldots, \mathbf{u}_m) = \mathcal{M}$, and
2. $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are linear independent.

- That is, a basis is a set of vectors that spans a linear subspace $\mathcal{M}$ without redundancy.
- An orthonormal basis (ONB), if in addition $\mathbf{u}_j$'s are unit vectors and orthogonal to each other.

- If $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a basis for a subspace $\mathcal{M}$, then any vector in $\mathcal{M}$ can be uniquely represented by the linear combination of $\mathbf{u}_i$'s. That is, if we can write a vector $\mathbf{v}$ as

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m, \quad \text{and also} \\ \mathbf{v} &= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_m\mathbf{u}_m, \end{aligned}$$

  then $c_i = a_i$ for all $i = 1 : m$.

- Bases are not unique. That is, a linear space $\mathcal{M}$ has more than one bases, but the number of vectors in each basis is the same, which is the rank/dim of $\mathcal{M}$.

- If $\mathcal{M}$ has dim p, then any set of more than p vectors from $\mathcal{M}$ is linear dependent.

- If $\mathcal{M}$ has dim p, then any set of p linearly independent vectors in $\mathcal{M}$ form a basis for $\mathcal{M}$.

Gram-Schmidt Procedure

Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$, we can use the *Gram-Schmidt* algorithm to obtain an ONB.

- $\mathbf{u}_1 = \mathbf{x}_1, \quad \mathbf{e}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$
- $\mathbf{u}_2 = \mathbf{x}_2 - (\mathbf{x}_2^t \mathbf{e}_1)\mathbf{e}_1, \quad \mathbf{e}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}$
- $\cdots$
- $\mathbf{u}_{k+1} = \mathbf{x}_{k+1} - \sum_{j=1}^{k} (\mathbf{x}_j^t \mathbf{e}_j)\mathbf{e}_j, \quad \mathbf{e}_{k+1} = \frac{\mathbf{u}_{k+1}}{\|\mathbf{u}_{k+1}\|}$

The resulting ONB is $\{\mathbf{e}_1, \ldots, \mathbf{e}_p\}$.

OLS Solution

- Consider a linear model

$$y_i = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + err_i, \quad i = 1, \ldots, n$$

Using the LS principal, we aim to find $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^t$, which minimizes

$$\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - \cdots - x_{ip}\beta_p)^2.$$

- Using the matrix form, we can write the linear model as

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \mathbf{e},$$

and solve

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2. \qquad (1)$$

The LS optimization

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2.$$

is equivalent to finding a vector $\mathbf{z}^*$ in $C(\mathbf{X})$, such that

$$\|\mathbf{y} - \mathbf{z}^*\| = \min_{\mathbf{z} \in C(\mathbf{X})} \|\mathbf{y} - \mathbf{z}\|^2.$$

Once we solve $\mathbf{z}^*$, we then go back to find its representation $\boldsymbol{\beta}$.