Vectors

Vectors and the scalar multiplication and vector addition operations:

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

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, $\cos(\theta) = 0$, 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

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

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

- Consider a matrix $\mathbf{X}_{n\times p} = (\mathbf{x}_1 \mid \cdots \mid \mathbf{x}_p)$, where the j-th column $\mathbf{x}_j$ is a $n \times 1$ vector.

  All the linear combinations of the p columns are denoted by $C(\mathbf{X})$, i.e.,

$$C(\mathbf{X}) = \text{ All linear combinations of } \mathbf{x}_1, \ldots, \mathbf{x}_p.$$

- 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

- $C(\mathbf{X})$ forms a linear subspace: all items from $C(\mathbf{X})$ are vectors from $\mathbb{R}^n$, and

$$\text{if } \mathbf{u}, \mathbf{v} \in C(\mathbf{X}), \text{ then } a\mathbf{u} + b\mathbf{v} \in C(\mathbf{X}),$$

  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).

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

$$x_1 = \frac{1}{a_1}\big(\mathbf{z}_1 - a_2\mathbf{x}_2 - \cdots - a_p\mathbf{x}_p\big).$$

  That is, any linear combination of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)$ can be rewritten as a linear combination of $(\mathbf{z}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$.

- Let $\tilde{\mathbf{X}}$ be a 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 $\mathbf{u}^t\mathbf{v} = 0$.

  For example, $\hat{\mathbf{y}} \perp \mathbf{y} - \hat{\mathbf{y}}$.

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

  For example, $(\mathbf{y} - \hat{\mathbf{y}}) \perp C(\mathbf{X})$.

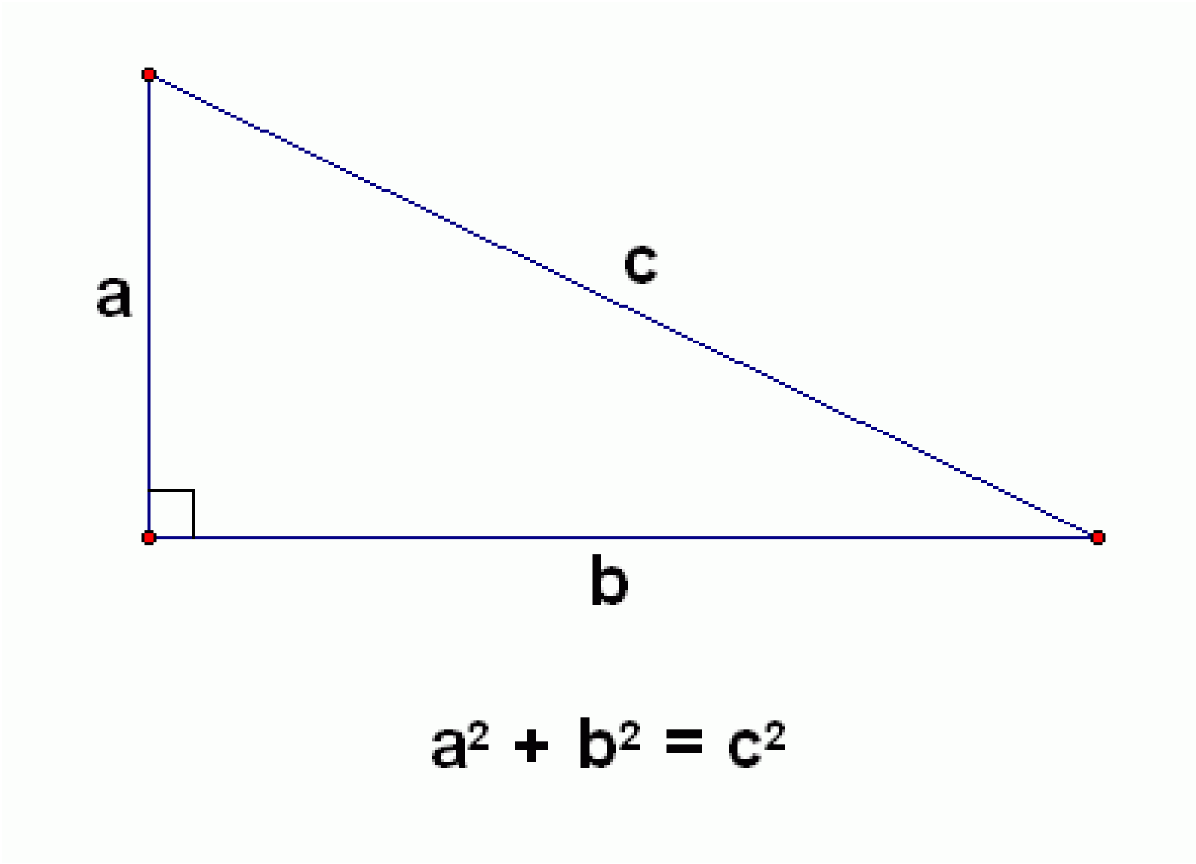
- Subspace $\perp$ Subspace: Similarly we can define orthogonal subspaces, if any vector from one subspace is orthogonal to any vector from the other subspace.

Pythagorean Theorem

If $\mathbf{v}_1 \perp \mathbf{v}_2$ then $\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2$.

In particular

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\mathbf{y} - \hat{\mathbf{y}}\|^2$$



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 **at least 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.
- $\mathbf{X}_{n\times p}$ is not of full rank $\iff$ its columns are linear dependent.
- $\mathbf{X}_{n\times p}$ is of full rank $\iff$ its columns are linear independent and form a basis for $C(\mathbf{X})$.

- $\mathbf{X} = (\mathbf{x}_1 \mid \cdots \mid \mathbf{x}_p)_{n\times p}$ is of full rank, then the p columns form a basis for $C(\mathbf{X})$. Any vector $\mathbf{v}$ in $C(\mathbf{X})$ can be uniquely represented by the linear combination of $\mathbf{x}_i$'s. That is, if we can write $\mathbf{v}$ as

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_p\mathbf{x}_p, \quad \text{and also} \\ \mathbf{v} &= a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_p\mathbf{x}_p, \end{aligned}$$

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

- **Bases are not unique.** That is, a linear space $C(\mathbf{X})$ has more than one bases, e.g., based on the replacement rule, we can replace $\mathbf{x}_j$ by another vector. But the number of vectors in each basis is always p, which is the rank/dim of $C(\mathbf{X})$.

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

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

The LS optimization

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

is equivalent to finding a vector $\mathbf{v}$ in $C(\mathbf{X})$ that minimizes $\|\mathbf{y}-\mathbf{v}\|^2$,

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

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

The optimal choice of $\mathbf{v}$ is $\hat{\mathbf{y}}$, the projection of $\mathbf{y}$ onto $C(\mathbf{X})$, the subspace consisting of linear combinations of columns of $\mathbf{X}$.

Projection

For any vector $\mathbf{y} \in \mathbb{R}^n$ and a subspace $\mathcal{M} \subseteq \mathbb{R}^n$, there exists a unique vector $\hat{\mathbf{y}}$ such that

1. $\hat{\mathbf{y}} \in \mathcal{M}$, and
2. $(\mathbf{y} - \hat{\mathbf{y}}) \perp \mathcal{M}$.

We call $\hat{\mathbf{y}}$ the **projection** of $\mathbf{y}$ onto $\mathcal{M}$.

$$\mathbf{y} = \underbrace{\hat{\mathbf{y}}}_{\in \mathcal{M}} + \big(\underbrace{\mathbf{y} - \hat{\mathbf{y}}}_{\in \mathcal{M}^\perp}\big)$$

The projection $\hat{\mathbf{y}}$ can be computed based on a set of basis of $\mathcal{M}$. More specifically,

$$\hat{\mathbf{y}}_{n\times 1} = \mathbf{M}_{n\times n}\mathbf{y}_{n\times 1}$$

where the $n \times n$ matrix $\mathbf{M}$ (known as the projection matrix) only depends on the underline subspace $\mathcal{M}$ and does not depend on $\mathbf{y}$. That is for any vector $\mathbf{y}$, we can compute $\mathbf{My}$ to obtain its projection.

LS and Projection

Recall the LS problem: find a vector $\mathbf{v}$ in $C(\mathbf{X})$, which minimizes $\|\mathbf{y}-\mathbf{v}\|^2$, i.e.,

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

Let $\mathbf{y}$ denote the projection of $\mathbf{y}$ onto $C(\mathbf{X})$. We have

$$\|\mathbf{y}-\mathbf{v}\|^2 = \| \underbrace{\mathbf{y}-\hat{\mathbf{y}}}_{\text{orthogonal to } C(\mathbf{X})} + \underbrace{\hat{\mathbf{y}}-\mathbf{v}}_{\in C(\mathbf{X})} \|^2 = \|\mathbf{y}-\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}}-\mathbf{v}\|^2 \geq \|\mathbf{y}-\hat{\mathbf{y}}\|^2.$$

So the LS solution is the projection of $\mathbf{y}$ onto the space $C(\mathbf{X})$:

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} = \mathbf{H}\mathbf{y}.$$

The projection matrix $\mathbf{H}$ is also called the hat matrix in many textbooks.

- If we apply some linear transformation on the columns of X, as long as $C(\mathbf{X})$ stays the same, $\hat{\mathbf{y}}$ and R^2 stay the same, although $\hat{\boldsymbol{\beta}}$ may differ.

- We can still compute $\hat{\mathbf{y}}$ even if $\mathbf{X}$ does not have full rank.

- $C(\mathbf{X})$ is often called the estimation space, and the residual vector $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{M})\mathbf{y}$ is orthogonal to $C(\mathbf{X})$, i.e., orthogonal to any linear combinations based on vectors from $C(\mathbf{X})$.

- The essence of LS: decompose the data vector $\mathbf{y}$ into two orthogonal components

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{r},$$

where $\hat{\mathbf{y}}$ in the estimation space and $\mathbf{r}$ in the error space.