Chapter 6: Dot product and orthogonality

San Jose State University

Prof. Guangliang Chen

Fall 2022

Outline

Introduction

In this lecutre we introduce geometric concepts such as

- length,
- distance,
- angle, and
- orthogonality

for vectors in $\mathbb{R}^n$.

They are all based on the so-called inner/dot product between vectors.

Dot product

Def 0.1. The **dot product**, also called **inner product**, between any two vectors of $\mathbb{R}^n$

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

is defined as

$$\underbrace{\mathbf{u} \cdot \mathbf{v}}_{\text{vector dot product}} = u_1 v_1 + \cdots + u_n v_n = [u_1 \ldots u_n] \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \underbrace{\mathbf{u}^T \mathbf{v}}_{\text{matrix product}}$$

Properties of the inner product

Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$, and c a scalar. Then

- $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.
- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
  $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
- $(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$

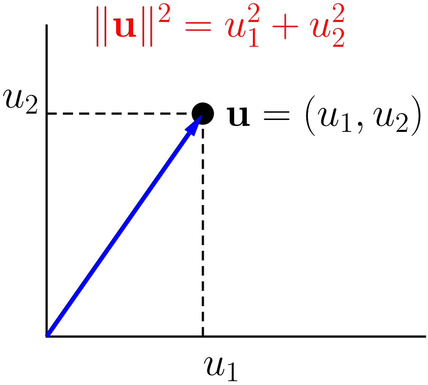
The length of a vector

Def 0.2. The **length** (or **norm**) of a vector of $\mathbb{R}^n$

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

is defined as

$$\|\mathbf{u}\| = \sqrt{\mathbf{u}\cdot\mathbf{u}} = \sqrt{u_1^2 + \cdots + u_n^2}$$



Properties of vector norm

Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$, and c a scalar. Then

- $\|\mathbf{u}\| \geq 0$ and $\|\mathbf{u}\| = 0$ if and only if $\mathbf{u} = \mathbf{0}$.
- $\|c\mathbf{u}\| = |c| \cdot \|\mathbf{u}\|$. In particular, $\|-\mathbf{u}\| = \|\mathbf{u}\|$.
- $\|\mathbf{u} \pm \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \pm 2\mathbf{u} \cdot \mathbf{v}$.
  This implies that $\|\mathbf{u} \pm \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

Proof.

- This is obviously true based on the definition.
- $\|c\mathbf{u}\| = \sqrt{(cu_1)^2 + \cdots + (cu_n)^2} = \sqrt{c^2(u_1^2 + \cdots + u_n^2)} = |c| \cdot \|\mathbf{u}\|$.
- We show the formula for $\mathbf{u} + \mathbf{v}$ first:

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} \\ &= \|\mathbf{u}\|^2 + 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{v}\|^2 \end{aligned}$$

  Now, apply this formula with $\mathbf{u}$ and $-\mathbf{v}$ gives the other formula:

$$\|\mathbf{u} + (-\mathbf{v})\|^2 = \|\mathbf{u}\|^2 + 2\mathbf{u} \cdot (-\mathbf{v}) + \| - \mathbf{v}\|^2$$

Unit vectors in $\mathbb{R}^n$

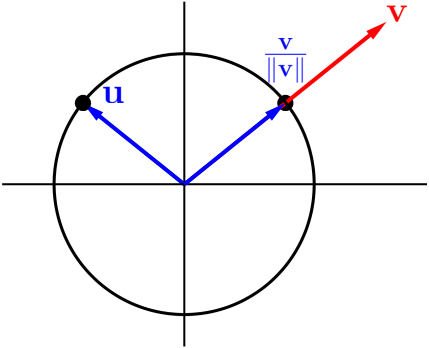
Def 0.3. A vector $\mathbf{u} \in \mathbb{R}^n$ whose length is 1 is called a **unit vector**.

Theorem 0.1. For any nonzero vector $\mathbf{v} \in \mathbb{R}^n$, the normalized form

$$\frac{1}{\|\mathbf{v}\|}\mathbf{v}$$

is a unit vector.

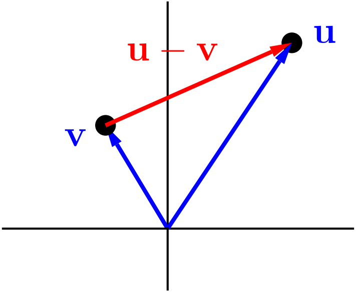
Proof. $\|\frac{1}{\|\mathbf{v}\|}\mathbf{v}\| = \frac{1}{\|\mathbf{v}\|} \cdot \|\mathbf{v}\| = 1.$ □



Distance in $\mathbb{R}^n$

Def 0.4. The distance between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is defined as

$$\begin{aligned} \operatorname{dist}(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\| \\ &= \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2} \end{aligned}$$

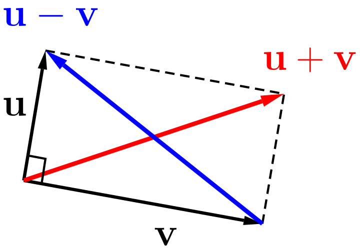


Orthogonal vectors

Def 0.5. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are said to be **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

Remark. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal if and only if

$$\|\mathbf{u} \pm \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

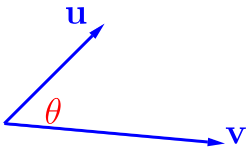


Angle between two vectors in $\mathbb{R}^n$

Def 0.6. The **angle** θ between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is defined as

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|} = \frac{\mathbf{u}}{\|\mathbf{u}\|}\cdot\frac{\mathbf{v}}{\|\mathbf{v}\|}$$

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$



Remark. Two special cases:

- $\mathbf{u}, \mathbf{v}$ are orthogonal ($\mathbf{u}\cdot\mathbf{v} = 0$): $\cos\theta = 0$ ($\theta = \frac{\pi}{2}$)
- $\mathbf{u}, \mathbf{v}$ coincide ($\mathbf{u} = \mathbf{v}$): $\cos\theta = 1$ ($\theta = 0$)

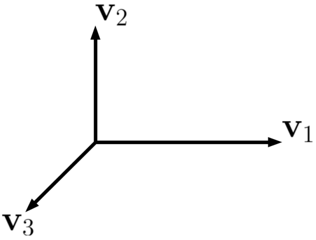
Example 0.1. Let $\mathbf{u} = [3, 4]^T, \mathbf{v} = [-1, 1]^T$. Compute the following:

- Dot product $\mathbf{u} \cdot \mathbf{v}$
- Norms of $\mathbf{u}, \frac{1}{5}\mathbf{u}, \mathbf{v}, -2\mathbf{v}$
- Distance between $\mathbf{u}, \mathbf{v}$
- Angle between $\mathbf{u}, \mathbf{v}$

Orthogonal sets

Def 0.7. A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is said to be an **orthogonal set** if each pair of vectors from the set is orthogonal, that is, if

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0, \quad \text{for all } i \neq j.$$



Example 0.2. The following sets of vectors of $\mathbb{R}^3$ are orthogonal sets:

- $\mathbf{e}_1 = [1, 0, 0]^T, \mathbf{e}_2 = [0, 1, 0]^T, \mathbf{e}_3 = [0, 0, 1]^T$
- $\mathbf{v}_1 = [1, 1, 1]^T, \mathbf{v}_2 = [1, -1, 0]^T, \mathbf{v}_3 = [1, 1, -2]^T$

Orthogonal sets must be linearly independent sets

Theorem 0.2. If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$ is an orthogonal set of nonzero vectors, then it is a linearly independent set.

Proof: Suppose

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

for some scalars $c_1, c_2, \ldots c_k$.

For each $i = 1, \ldots, k$, take dot product between $\mathbf{v}_i$ and each side of the equation to get

$$\mathbf{v}_i \cdot (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = \mathbf{v}_i \cdot \mathbf{0}$$

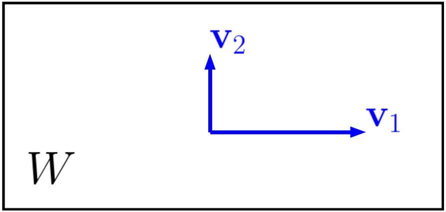
Since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are orthogonal to each other, we have

$$c_i(\mathbf{v}_i \cdot \mathbf{v}_i) \longleftarrow \mathbf{v}_i \cdot (c_i\mathbf{v}_i) = \mathbf{0}$$

Since $\mathbf{v}_i$ is nonzero, i.e., $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$, we obtain that $c_i = 0$. This thus completes the proof.

Orthogonal basis (basis + orthogonality)

Def 0.8. A basis $\mathcal{B}$ for a subspace W of $\mathbb{R}^n$ is called an **orthogonal basis** for W if $\mathcal{B}$ is also an orthogonal set.



Example 0.3. Each of the following two sets of vectors is an orthogonal basis for $\mathbb{R}^3$:

- $\mathbf{e}_1 = [1, 0, 0]^T, \mathbf{e}_2 = [0, 1, 0]^T, \mathbf{e}_3 = [0, 0, 1]^T$
- $\mathbf{v}_1 = [1, 1, 1]^T, \mathbf{v}_2 = [1, -1, 0]^T, \mathbf{v}_3 = [1, 1, -2]^T$

but the following sets are not:

- $\mathbf{v}_1 = [1, 1, 0]^T, \mathbf{v}_2 = [1, -1, 0]^T$ (only an orthogonal set)
- $\mathbf{v}_1 = [1, 0, 0]^T, \mathbf{v}_2 = [1, 1, 0]^T, \mathbf{v}_3 = [1, 1, 1]^T$ (only a basis)

Under an orthogonal basis, coordinates are easy to compute

Theorem 0.3. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For any vector $\mathbf{x} \in W$, the coordinate vector of $\mathbf{x}$ with respect to the basis is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}, \qquad \text{with} \quad c_i = \frac{\mathbf{x} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} = \frac{\mathbf{x} \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2}$$

This implies that

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \frac{\mathbf{x} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \cdots + \frac{\mathbf{x} \cdot \mathbf{v}_k}{\mathbf{v}_k \cdot \mathbf{v}_k}\mathbf{v}_k$$

Proof. Suppose

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{x}$$

for some scalars $c_1, c_2, \ldots c_k$. We need to solve for $c_1, \ldots, c_k$.

For each $i = 1, \ldots, k$, use $\mathbf{v}_i$ to take dot product with the equation to get

$$\begin{aligned} \mathbf{v}_i \cdot \mathbf{x} &= \mathbf{v}_i \cdot (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \\ &= \mathbf{v}_i \cdot (c_i\mathbf{v}_i) = c_i(\mathbf{v}_i \cdot \mathbf{v}_i) \end{aligned}$$

where we have used the orthogonality of the vectors.

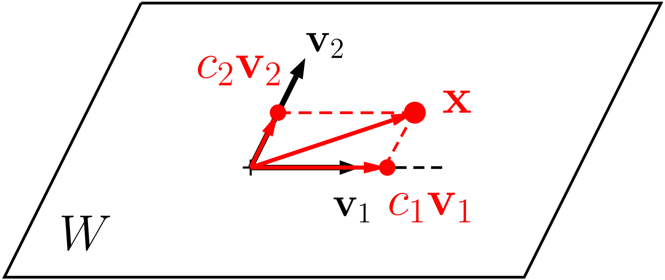
Since $\mathbf{v}_i$ is nonzero, i.e., $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$, we obtain that

$$c_i = \frac{\mathbf{v}_i \cdot \mathbf{x}}{\mathbf{v}_i \cdot \mathbf{v}_i}$$

This thus completes the proof.

Illustration: Coordinates relative to an orthogonal basis

$$(c_1, c_2) = \left(\frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}, \frac{\mathbf{x}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\right)$$

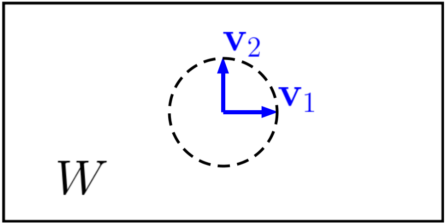


Example 0.4. For the coordinate vector of $\mathbf{x} = [1, 2, 3]^T$ with respect to the orthogonal basis

$$\mathbf{v}_1 = [1, 1, 1]^T, \mathbf{v}_2 = [1, -1, 0]^T, \mathbf{v}_3 = [1, 1, -2]^T$$

Orthonormal = orthogonal + unit length

- A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$ is called an **orthonormal set** if the vectors are orthogonal to each other and all have unit norm.
- An orthogonal basis for a subspace of $\mathbb{R}^n$ is called an **orthonormal basis** if the basis vectors all have unit norm.



Example 0.5. Each of the following sets of vectors is an orthonormal basis for $\mathbb{R}^3$:

- $\mathbf{e}_1 = [1, 0, 0]^T, \mathbf{e}_2 = [0, 1, 0]^T, \mathbf{e}_3 = [0, 0, 1]^T$
- $\mathbf{v}_1 = \frac{1}{\sqrt{3}}[1, 1, 1]^T, \mathbf{v}_2 = \frac{1}{\sqrt{2}}[1, -1, 0]^T, \mathbf{v}_3 = \frac{1}{\sqrt{6}}[1, 1, -2]^T$

Expansion onto an orthonormal basis is even easier

Corollary 0.4. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$. For any vector $\mathbf{x} \in W$, the coordinate vector of $\mathbf{x}$ with respect to the basis is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}, \qquad \text{with} \quad c_i = \mathbf{x} \cdot \mathbf{v}_i$$

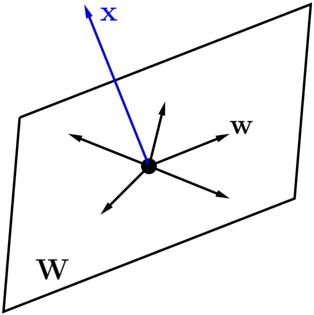
This implies that

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = (\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + \cdots + (\mathbf{x} \cdot \mathbf{v}_k)\mathbf{v}_k$$

Example 0.6. Find the coordinates of $\mathbf{x} = [1, 2, 3]^T$ with respect to the orthonormal basis $\mathbf{v}_1 = \frac{1}{\sqrt{3}}[1, 1, 1]^T, \mathbf{v}_2 = \frac{1}{\sqrt{2}}[1, -1, 0]^T, \mathbf{v}_3 = \frac{1}{\sqrt{6}}[1, 1, -2]^T$

Orthogonal subspaces

Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ a vector in $\mathbb{R}^n$. We say that $\mathbf{x}$ is **orthogonal** to W if $\mathbf{x} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in W$, and denote it by $\mathbf{x} \perp W$.



Def 0.9. Let U, V be two subspaces of $\mathbb{R}^n$.

- The two subspaces U, V are said to be **orthogonal** to each other if every vector $\mathbf{u} \in U$ is orthogonal to V and every vector $\mathbf{v} \in V$ is orthogonal to U. That is,

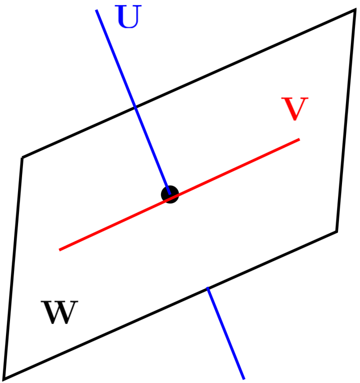
$$\mathbf{u} \cdot \mathbf{v} = 0, \quad \text{for all } \mathbf{u} \in U, \mathbf{v} \in V.$$

- They are called **orthogonal complements** of each other in $\mathbb{R}^n$ if they are orthogonal to each other and their total dimension is equal to n (i.e., $\dim(U) + \dim(V) = n$). In this case, we write $U = V^\perp$ and $V = U^\perp$.

Example 0.7. In the right picture, U, V, W are all subspaces of $\mathbb{R}^3$.

- **orthogonal subsapces**: U and V, U and W
- **orthogonal complements**: only U and W.

  We thus write $U = W^\perp$ and $W = U^\perp$.



$\mathrm{Row}(\mathbf{A}), \mathrm{Nul}(\mathbf{A})$ are orthogonal complements

For any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, one can define three kinds of susbpaces, but only two of them belong to the same vector space:

$$\mathrm{Row}(\mathbf{A}),\ \mathrm{Nul}(\mathbf{A}) \subseteq \mathbb{R}^n \qquad (\mathrm{Col}(\mathbf{A}) \subseteq \mathbb{R}^m)$$

In fact, these two must be orthogonal complements.

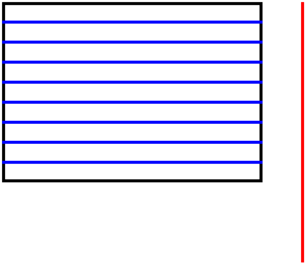
Theorem 0.5. For any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$,

$$(\mathrm{Row}(\mathbf{A}))^{\perp} = \mathrm{Nul}(\mathbf{A})$$

Dot product and orthogonality

To prove that $\text{Row}(\mathbf{A}),\ \text{Nul}(\mathbf{A})$ are orthogonal complements, we need to verify

(1) The two subspaces are orthogonal to each other:



$= \mathbf{0}$

(2) Their total dimension is n, i.e., $\dim(\text{Row}(\mathbf{A})) + \dim(\text{Nul}(\mathbf{A})) = n$. This is because

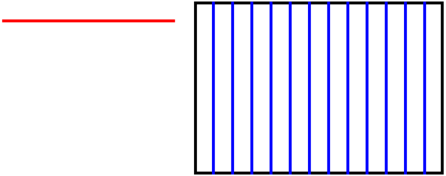
- $\dim(\text{Row}(\mathbf{A})) = \text{rank}(\mathbf{A}) =$ #pivots;
- $\dim(\text{Nul}(\mathbf{A})) = n - \text{rank}(\mathbf{A}) =$ #free variables

Remark. The theorem implies that the orthogonal complement of $\mathrm{Col}(\mathbf{A})$ in $\mathbb{R}^m$ is $\mathrm{Nul}(\mathbf{A}^T)$:

$$(\mathrm{Col}(\mathbf{A}))^\perp = \left(\mathrm{Row}(\mathbf{A}^T)\right)^\perp = \mathrm{Nul}(\mathbf{A}^T)$$

where

$$\mathrm{Nul}(\mathbf{A}^T) = \{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{A}^T\mathbf{x} = \mathbf{0}\} = \{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{x}^T\mathbf{A} = \mathbf{0}^T\}$$



$= \mathbf{0}^T$

Example 0.8. Consider the following matrix and its RREF

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

We have

- $\text{Row}(\mathbf{A}) = \text{span}\{[1, 0, -1]^T, [0, 1, 2]^T\}$, and
- $\text{Nul}(\mathbf{A}) = \text{span}\{[1, -2, 1]^T\}$.

The two subspaces are orthogonal complements of each other (inside $\mathbb{R}^3$).

On the other hand, $\text{Col}(\mathbf{A}) = \mathbb{R}^2$ and $\text{Nul}(\mathbf{A}^T) = \{\mathbf{0}\}$. The two subspaces are also orthogonal complements of each other (in $\mathbb{R}^2$).

Orthogonal matrix (square matrix w/ orthonormal columns)

Def 0.10. A square matrix $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n] \in \mathbb{R}^{n\times n}$ is called an **orthogonal** matrix if its columns are an orthonormal set of vectors, i.e.,

$$\mathbf{q}_i \cdot \mathbf{q}_j = \begin{cases} 1, & i = j \longleftarrow \text{Unit norm} \\ 0, & i \neq j \longleftarrow \text{Orthogonality} \end{cases}$$

Remark. The columns of an $n \times n$ orthogonal matrix must form an orthonormal basis for $\mathbb{R}^n$ (and vice versa).

Example 0.9. The following is an example of an orthogonal matrix (because the columns of the matrix form an orthonormal basis for $\mathbb{R}^3$):

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{bmatrix}$$

Inverse of an orthogonal matrix is its transpose

Theorem 0.6. If $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix, then $\mathbf{Q}^{-1} = \mathbf{Q}^T$. The converse is also true.

Proof.

$$\mathbf{Q}^T\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} [\mathbf{q}_1\,\mathbf{q}_2 \ldots \mathbf{q}_n] = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = \mathbf{I}_n$$

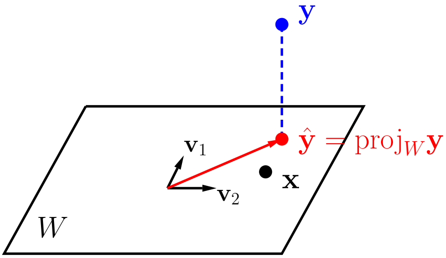
□

The orthogonal projection problem

Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. We have showed that if $\mathbf{x}$ lies in W, then it can be represented as

$$\mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \cdots + \frac{\mathbf{x}\cdot\mathbf{v}_k}{\mathbf{v}_k\cdot\mathbf{v}_k}\mathbf{v}_k$$

For any vector $\mathbf{y}$ outside of W, its orthogonal projection onto W, $\hat{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$, will be inside W.



It can be shown that $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$.

Question: How can we find $\hat{\mathbf{y}}$?

Remark. In order for $\hat{\mathbf{y}}$ to be the orthogonal projection of $\mathbf{y}$ onto W, we must have

$$\mathbf{y} - \hat{\mathbf{y}} \perp W.$$

In particular,

$$\mathbf{y} - \hat{\mathbf{y}} \perp \mathbf{v}_i,\ 1 \le i \le k \qquad \text{and} \quad \mathbf{y} - \hat{\mathbf{y}} \perp \hat{\mathbf{y}}.$$

This also leads to a decomposition of $\mathbf{y}$ along W and $W^\perp$:

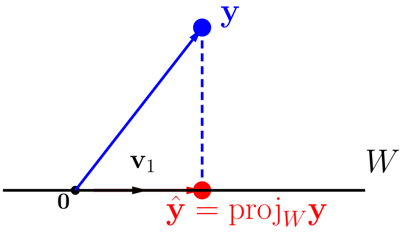
$$\mathbf{y} = \underbrace{\hat{\mathbf{y}}}_{\in W} + \underbrace{(\mathbf{y} - \hat{\mathbf{y}})}_{\in W^\perp}$$

Lastly, the distance from $\mathbf{y}$ to W can be defined as follows:

$$\mathrm{dist}(\mathbf{y}, W) = \|\mathbf{y} - \hat{\mathbf{y}}\|, \quad \hat{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$$

The case of $k = 1$

We first consider the projection of a point onto a 1-dimensional subspace spanned by a single vector $\mathbf{v}_1$.



Suppose $\hat{\mathbf{y}} = c_1\mathbf{v}_1$ (with c_1 TBD). Since $\mathbf{y} - \hat{\mathbf{y}}$ must be orthogonal to W, we have

$$\begin{aligned} 0 &= \mathbf{v}_1 \cdot (\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{v}_1 \cdot (\mathbf{y} - c_1\mathbf{v}_1) \\ &= \mathbf{v}_1 \cdot \mathbf{y} - c_1 \mathbf{v}_1 \cdot \mathbf{v}_1 \end{aligned}$$

This yields that

$$c_1 = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \longrightarrow \hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

Example 0.10. Let $\mathbf{v} = [3, 4]^T$. Find the projection of $\mathbf{x} = [1, 0]^T$ onto the subspace spanned by $\mathbf{v}$. What is the distance from $\mathbf{x}$ to the subspace?

The case of $k = 2$

When $k = 2$, suppose $\hat{\mathbf{y}} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ (with c_1, c_2 TBD).

Since $\mathbf{y} - \hat{\mathbf{y}}$ must be orthogonal to W, and in particular, $\mathbf{y} - \hat{\mathbf{y}}$ must be orthogonal to $\mathbf{v}_1$, we have

$$\begin{aligned} 0 &= \mathbf{v}_1 \cdot (\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbf{v}_1 \cdot (\mathbf{y} - c_1\mathbf{v}_1 - c_2\mathbf{v}_2) \\ &= \mathbf{v}_1 \cdot \mathbf{y} - c_1\mathbf{v}_1 \cdot \mathbf{v}_1 - 0 \end{aligned}$$

from which we obtain that

$$c_1 = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}$$

Similarly, $\mathbf{y} - \hat{\mathbf{y}}$ must be orthogonal to $\mathbf{v}_2$:

$$0 = \mathbf{v}_2 \cdot (\mathbf{y} - \hat{\mathbf{y}})$$

From this, we obtain that

$$c_2 = \frac{\mathbf{y} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}$$

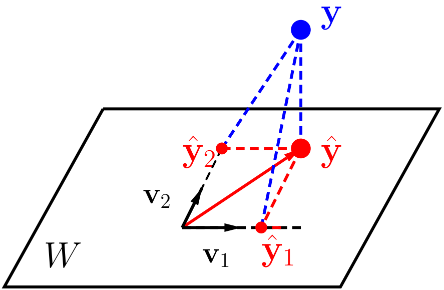
Putting everything together,

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{y} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

Geometric interpretation

Projection onto a subspace (with an orthogonal basis) is equal to the sum of projections onto the basis vectors individually:

$$\hat{\mathbf{y}} = \underbrace{\frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1}_{\hat{\mathbf{y}}_1} + \underbrace{\frac{\mathbf{y}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2}_{\hat{\mathbf{y}}_2}$$



Example 0.11. Let $\mathbf{v}_1 = [1, 1, 0]^T, \mathbf{v}_2 = [1, -1, 0]^T$. Find the projection of $\mathbf{x} = [2, 3, 4]^T$ onto the subspace spanned by $\mathbf{v}_1, \mathbf{v}_2$.

The general case of any k

The previous approach applies to any k, leading the following result.

Theorem 0.7. The orthogonal projection of any vector $\mathbf{y} \in \mathbb{R}^n$ onto a subspace W, with an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, is

$$\mathrm{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{v}_k}{\mathbf{v}_k \cdot \mathbf{v}_k}\mathbf{v}_k$$

Remark. If the orthogonal basis is orthonormal, then the formula simplifies to

$$\mathrm{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{v}_1)\mathbf{v}_1 + \cdots + (\mathbf{y} \cdot \mathbf{v}_k)\mathbf{v}_k$$

The Gram-Schmidt Orthogonalization Process

Orthogonal bases are great because they simplify the math in many cases, such as finding coordinate vectors and orthogonal projections.

An important question would be, how do we construct orthogonal bases?

The Gram-Schmidt process is a procedure that converts any given basis of a subspace to an orthogonal basis for the same subspace:

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \text{ (general basis)} \quad \longrightarrow \quad \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \text{ (orthogonal basis)}$$

Theorem 0.8 (Gram-Schmidt Orthogonalization). Given a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for a nonzero subspace $W \subseteq \mathbb{R}^n$, the following vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ form an orthogonal basis for W:

$$\mathbf{u}_1 = \mathbf{v}_1$$

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

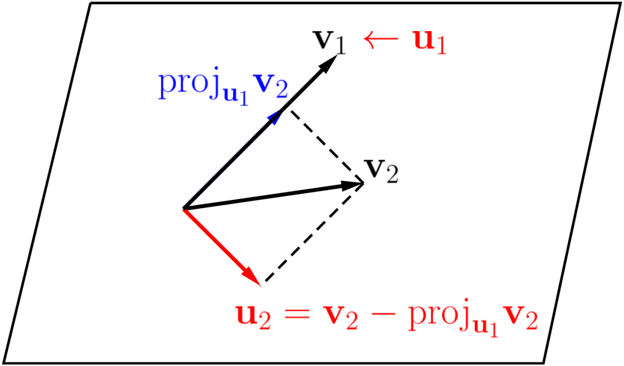
$$\mathbf{u}_3 = \mathbf{v}_3 - \text{proj}_{\mathbf{u}_1,\mathbf{u}_2}\mathbf{v}_3 = \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2$$

$$\cdots$$

$$\mathbf{u}_k = \mathbf{v}_k - \text{proj}_{\mathbf{u}_1,\ldots,\mathbf{u}_{k-1}}\mathbf{v}_k = \mathbf{v}_k - \frac{\mathbf{v}_k \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 - \cdots - \frac{\mathbf{v}_k \cdot \mathbf{u}_{k-1}}{\mathbf{u}_{k-1} \cdot \mathbf{u}_{k-1}}\mathbf{u}_{k-1}$$

Remark. To further get an orthonormal basis, just normalize each $\mathbf{u}_i$.

Gram Schmidt: $\{\mathbf{v}_1, \mathbf{v}_2\} \longrightarrow \{\mathbf{u}_1, \mathbf{u}_2\}$



Example 0.12. Given a basis for $\mathbb{R}^2$: $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, construct an orthogonal basis from it.

Example 0.13. Find an orthogonal basis for the span of

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Example 0.14. Find an orthogonal basis for the span of

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 5 \end{bmatrix}.$$

How can we further obtain an orthonormal basis?

Least-squares (LS) problems

Often we encounter inconsistent systems of linear equations (e.g., due to contradictions among the equations):

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

Though an exact solution does not exist, we can still hope to find an $\mathbf{x}$ such that $\mathbf{A}\mathbf{x}$ is as close to $\mathbf{b}$ as possible, i.e.,

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}$$

The specify what we mean by “close”, we need to choose a criterion. Then the solution is said to be optimal under the chosen criterion.

For example, the following system has no exact solution:

$$\begin{cases} x + y = 3 \\ x - y = 1 \\ 2x + 3y = 6.4 \end{cases} \qquad \longrightarrow \qquad \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 6.4 \end{bmatrix}$$

but the pair of $x = 1.92, y = 0.88$ makes all equations nearly true $(2.8, 1.04, 6.48)$.

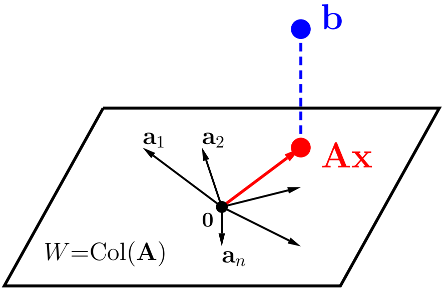
We shall see that this solution is optimal under the so-called least squares criterion.

Mathematical formulation of LS problems

Formally, we formulate the following LS problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|$$

where $\mathbf{A} = [\mathbf{a}_1\, \mathbf{a}_2 \ldots \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ (with $m \geq n$), and $\mathbf{b} \in \mathbb{R}^m$ are both given.



The solution to the above LS problem is called the **LS solution** of the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Since $\mathbf{A}\mathbf{x} \in \mathrm{Col}(\mathbf{A})$ for all $\mathbf{x}$, we are looking for the closest vector to $\mathbf{b}$ in the column space of $\mathbf{A}$.

The least squares solution $\mathbf{x}$ should be such that $\mathbf{b} - \mathbf{A}\mathbf{x} \perp \mathrm{Col}(\mathbf{A})$. In particular, it is orthogonal to every column of $\mathbf{A}$:

$$\mathbf{a}_1^T(\mathbf{b} - \mathbf{A}\mathbf{x}) = 0, \quad \mathbf{a}_2^T(\mathbf{b} - \mathbf{A}\mathbf{x}) = 0, \quad \ldots, \quad \mathbf{a}_n^T(\mathbf{b} - \mathbf{A}\mathbf{x}) = 0$$

These equations can be combined together as follows:

$$\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}) = \mathbf{0} \;\longrightarrow\; \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

This equation has a unique solution when $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ is invertible:

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}.$$

Remark. The invertibility condition for $\mathbf{A}^T\mathbf{A}$ holds true if and only if all the columns of $\mathbf{A}$ are linearly independent, in which case we say that $\mathbf{A}$ is of full column rank.

The reason is that for any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$

$$\mathrm{rank}(\mathbf{A}^T\mathbf{A}) = \mathrm{rank}(\mathbf{A}).$$

We prove this result by showing that the two matrices have the same null space, i.e., $\mathrm{Nul}(\mathbf{A}^T\mathbf{A}) = \mathrm{Nul}(\mathbf{A})$.

Proof of $\mathrm{Nul}(\mathbf{A}^T\mathbf{A}) = \mathrm{Nul}(\mathbf{A})$:

(1) $\mathrm{Nul}(\mathbf{A}) \subseteq \mathrm{Nul}(\mathbf{A}^T\mathbf{A})$: Suppose that $\mathbf{x} \in \mathrm{Nul}(\mathbf{A})$, i.e., $\mathbf{A}\mathbf{x} = \mathbf{0}$. Multiplying both sides by $\mathbf{A}^T$ gives that $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{0} = \mathbf{0}$. This shows that $\mathbf{x} \in \mathrm{Nul}(\mathbf{A}^T\mathbf{A})$.

(2) $\mathrm{Nul}(\mathbf{A}^T\mathbf{A}) \subseteq \mathrm{Nul}(\mathbf{A})$: Suppose that $\mathbf{x} \in \mathrm{Nul}(\mathbf{A}^T\mathbf{A})$, i.e., $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{0}$. Multiplying both sides by $\mathbf{x}^T$ gives that

$$\mathbf{0} = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2.$$

This implies that $\mathbf{A}\mathbf{x} = \mathbf{0}$ and thus that $\mathbf{x} \in \mathrm{Nul}(\mathbf{A})$.

We have thus obtained the following result.

Theorem 0.9. If $\mathbf{A}$ is of full column rank (i.e., it has linearly independent columns), then the following problem

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|$$

has a unique solution

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}.$$

Remark. The LS approximation error is

$$\| \underbrace{\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T}_{\mathbf{H}\in\mathbb{R}^{n\times n},\,\text{Hat matrix}} \mathbf{b} - \mathbf{b}\|$$

Example 0.15. Verify that the least squares solution of the following linear system

$$\begin{cases} x + y = 3 \\ x - y = 1 \\ 2x + 3y = 6.4 \end{cases}$$

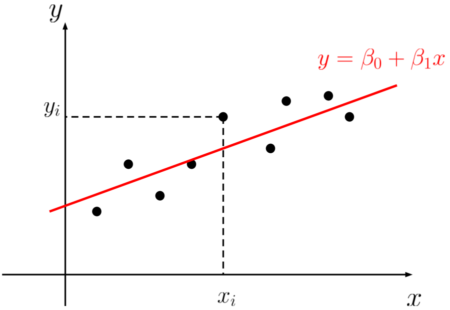
is $x = 1.92, y = 0.88$.

Application to simple linear regression

Given data $(x_1, y_1), \ldots, (x_n, y_n)$, we would like to fit a line $y = \beta_0 + \beta_1 x$ (exactly or as closely as possible to the data):

$$\beta_0 + \beta_1 x_i = y_i, \quad i = 1, \ldots, n$$

This is a linear system consisting of n equations, in two unknowns (β_0, β_1). It typically has no exact solution due to noise.



We can derive the matrix equation corresponding to the above problem, as well as its LS solution.

Let

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Then the linear system can be written as

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$$

The least squares solution is given by

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

It follows that the LS regression line is given by

$$y = \beta_0 + \beta_1 x$$

where β_0, β_1 are the components of $\boldsymbol{\beta}$.

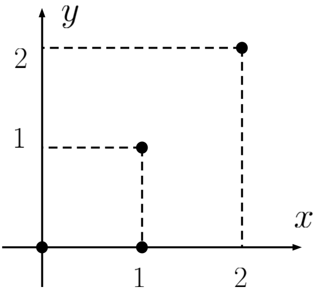
Remark. The LS fitted values are

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

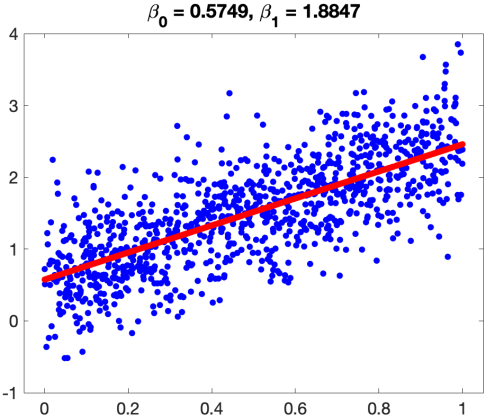
and the LS fitting error is

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\| = \left\|\left(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)\mathbf{y}\right\|$$

Example 0.16. Given the following data, find the least-squares regression line. What are the LS fitted values and total fitting error?



Remark. To perform linear regression on larger data sets, use software.



The multiple linear regression problem

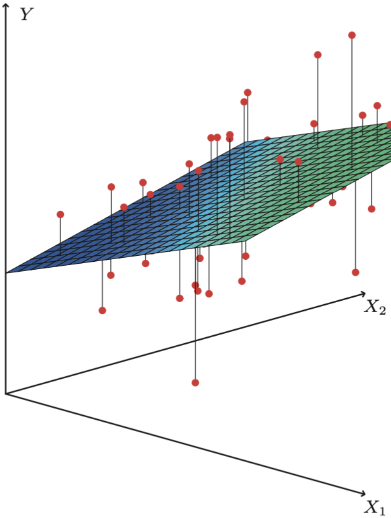
Consider a linear model with multiple predictors

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

where

- y: response,
- $x_1, \ldots, x_k$: predictors
- $\beta_0, \beta_1, \ldots, \beta_k$: coefficients (unknown)

Dot product and orthogonality



Assume n observations of the response and predictors (subject to noise),

$$(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i), \quad 1 \le i \le n$$

We would like to use the data to estimate $\beta_0, \beta_1, \ldots, \beta_k$ such that

$$y_i \approx \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}, \quad 1 \le i \le n$$

Let $p = k + 1$ (#regression coefficients including the intercept) and

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

The goal of multiple linear regression is to estimate $\boldsymbol{\beta}$ such that

$$\underbrace{\mathbf{y}}_{n\times 1} \approx \underbrace{\mathbf{X}}_{n\times p} \cdot \underbrace{\boldsymbol{\beta}}_{p\times 1}$$

Under the LS criterion, the regression coefficients can be found by using symbolically the same formula.

Theorem 0.10. If $\mathbf{X}$ is of full column rank, then the LS solution of the multiple linear regression problem is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$