Matrix Algebra and Applications

MATH 2111 Lecture Notes

LI, Hongrui

December 5, 2024

Contents

6 Inner Product Spaces

6.1 Inner Product, Length, and Orthogonality

In $\mathbb{R}^2, \mathbb{R}^3$, we have already learnt the concepts of dot product, length/distance, perpendicularity, etc. In this section, we will generalize these concepts to $\mathbb{R}^n$ and more general vector spaces.

- **Inner Product:** Let $\mathbf{u} = (u_1, u_2, \cdots, u_n)$ and $\mathbf{v} = (v_1, v_2, \cdots, v_n)$ be vectors in $\mathbb{R}^n$. The **inner product** of $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

- **Length:** The **length** (or **norm**) of a vector $\mathbf{v}$ in $\mathbb{R}^n$ is defined as

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

- **Distance:** The **distance** between two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is defined as

$$d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}.$$

- **Orthogonality:** Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

> **Proposition 6.1. Pythagorean Theorem**
> Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if and only if
>
> $$||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2.$$

Orthogonal Complements

Let $\mathbf{z} \in \mathbb{R}^n$ and W is a subspace of $\mathbb{R}^n$. If $\mathbf{z}$ is orthogonal to every vector in W, we say that $\mathbf{z}$ is orthogonal to W, and write $\mathbf{z} \perp W$. The set of all vectors in $\mathbb{R}^n$ that are orthogonal to W is called the **orthogonal complement** of W, denoted by $W^\perp$. By definition, we can easily see that $W^\perp$ is also a subspace of $\mathbb{R}^n$. Furthermore, we will look at some facts:

- Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of A is the null space of A^T, that is

$$\text{Row} A^\perp = \text{Nul} A, \quad \text{Col} A^\perp = \text{Nul} A^T.$$

- Let W be a subspace of $\mathbb{R}^n$. Then $(W^\perp)$ is also a subspace of $\mathbb{R}^n$, and

$$\dim W + \dim W^\perp = n.$$

6.2 Orthogonal Sets

In applications in different areas, like the Cartesian coordinate system, we use a set of orthogonal vectors to represent the space, and use them as the basis to represent other instances like vectors, surfaces, etc. Here, we summarize such set of vectors as orthogonal sets and show some properties of them.

> **Definition 6.1.** A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distince vectors in the set is orthogonal, that is, $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for all $i \neq j$.

> **Proposition 6.2.** If $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then the set S is linearly independent and hence is a basis for the subspace V of $\mathbb{R}^n$ spanned by S.

Proof. Suppose that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$, then we have

$$\begin{aligned}
\mathbf{0} &= \mathbf{v}_1 \cdot (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) \\
&= c_1\mathbf{v}_1 \cdot \mathbf{v}_1 + c_2\mathbf{v}_1 \cdot \mathbf{v}_2 + \cdots + c_n\mathbf{v}_1 \cdot \mathbf{v}_n \\
&= c_1\|\mathbf{v}_1\|^2,
\end{aligned}$$

which implies that $c_1 = 0$. Similarly, we can show that $c_2 = c_3 = \cdots = c_n = 0$. Thus, by contradiction, we can show that the set S is linearly independent. $\square$

Definition 6.2. An **orthogonal basis** for a subspace V of $\mathbb{R}^n$ is a basis for V that is also an orthogonal set.

Proposition 6.3. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is an orthogonal basis for a subspace V of $\mathbb{R}^n$. For each $\mathbf{y} \in V$, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

are given by

$$c_i = \frac{\mathbf{y} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}, \quad i = 1, 2, \cdots, n.$$

The proof is trivial by multiplying $\mathbf{v}_i$ on both sides of the equation above, but always keep in mind that if the basis were not orthogonal, it would be necessary to solve a system of linear equations in order to find the weights.

In the 2D Cartesian coordinate system we have learnt before, we covered the concept of projection vectors. Here, we discuss the similar idea in the context of orthogonal sets and expand it to $\mathbb{R}^n$.

Definition 6.3. Let $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, and denote

$$L \equiv \text{span}\{\mathbf{u}\}, \text{ the line parralel to } \mathbf{u}.$$

Write $\mathbf{y}$ as
$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{w}$$

such that $\hat{\mathbf{y}} = c\mathbf{u}$ (parallel to $\mathbf{u}$) and $\mathbf{w}$ is orthogonal to $\mathbf{u}$. Then $\hat{\mathbf{y}}$ is called the **orthogonal projection** of $\mathbf{y}$ onto $\mathbf{u}$, denoted by $\text{proj}_\mathbf{u}\mathbf{y}$ or $\text{proj}_L\mathbf{y}$. and $\mathbf{w}$ is called the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$.

Just like the projection vector in 2D, we can calculate the orthogonal projection by the following formula:
$$\hat{\mathbf{y}} = \text{proj}_\mathbf{u}\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}, \quad \mathbf{w} = \mathbf{y} - \hat{\mathbf{y}}.$$

Recall that in the Cartesian coordinate system, we require the basis to be not only orthogonal, but also normalized. In the context of $\mathbb{R}^n$, we have the word of "orthonormality" to describe such basis.

Definition 6.4. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ in $\mathbb{R}^n$ is said to be an **orthonormal set** if it is an orthogonal set and each vector in the set has unit length, that is, $\|\mathbf{v}_i\| = 1$ for all i.
If W is a subspace spanned by an orthonormal set, then the set is called an **orthonormal basis** for W.

Proposition 6.4. An $n \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

The proof of this proposition follows directly from how we calculate the product of two matrices, and the readers can try to prove it by themselves. This reveals the relationship between orthonormal sets and square matrices, and next, we will discuss some properties of general $m \times n$ matrices.

Proposition 6.5. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^n$. Then

- $||U\mathbf{x}|| = ||\mathbf{x}||$

- $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$

- $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{0}$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$

Remark: Properties a and c say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality.

To bring this section to an end, we introduce the following lemmma of the propositions above, whose proof is left as an exercise to the readers.

Let U be an $n \times n$ matrix with orthonormal columns, show that $\det U = \pm 1$.

6.3 Orthogonal Projections

In the previous section, we have introduced the concept of orthogonal projection onto a vector, as well as the orthonormal set. Now let's combine them together and introduce the orthogonal projection onto a subspace.

Example 6.1. Let W to be the xy-plane, and $\mathbf{u_1} \in W, \mathbf{u_2} \in W^\perp$, then we can decomposit the vector $\mathbf{u} = (a, b, c)$ into two parts:

$$\mathbf{u} = (a, b, c) = (a, b, 0) + (0, 0, c) = \mathbf{u_1} + \mathbf{u_2}$$

In general, we want to find a decomposition of a vector $\mathbf{y}$ in $\mathbb{R}^n$ into two parts, one in the subspace W and the other in the orthogonal complement $W^\perp$.

Proposition 6.6. Orthogonal Decomposition Theorem
Let W be a subspace of $\mathbb{R}^n$, then each $\mathbf{y} \in \mathbb{R}^n$ can be uniquely written as

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In fact, if $\{\mathbf{u_1}, \cdots, \mathbf{u_p}\}$ is any orthogonal basis for W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u_1}}{\mathbf{u_1} \cdot \mathbf{u_1}} \mathbf{u_1} + \cdots + \frac{\mathbf{y} \cdot \mathbf{u_p}}{\mathbf{u_p} \cdot \mathbf{u_p}} \mathbf{u_p}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.
Here, $\hat{\mathbf{y}}$ is called the **orthogonal projection** of $\mathbf{y}$ onto W, denoted by $\text{proj}_W \mathbf{y}$.

We may inteprete the the orthogonal projection onto a orthogonal basis as a generalization of the orthogonal projection onto a vector. Hence, the proof of this propositon is similar that to the latter. Since the correctness of a decomposition can be easily verified, by checking whether $\mathbf{z}$ is orthogonal to each $\mathbf{u_i}$, we will not provide examples here. Readers can randomly choose a vector and a subspace as exercises and verify the correctness by themselves.

Proof. The proof of this proposition can be slipped into two parts, i.e., the existence and uniqueness, and here is an outline of the proof:

- **Existence:** First we observe that $\hat{\mathbf{y}}$ is in W, since it's a linear combination of the basis vectors, and next we only need to show that $\mathbf{z}$ is orthogonal to W. This can be done by calculating the dot product of $\mathbf{z}$ and $\mathbf{u_i}$, and we can show that $\mathbf{z}$ is orthogonal to each $\mathbf{u_i}$.

- **Uniqueness:** Suppose that there are two, then $\hat{\mathbf{y}} = \hat{\mathbf{y_1}} + \hat{\mathbf{z_1}} = \hat{\mathbf{y_2}} + \hat{\mathbf{z_2}}$, and

$$(\hat{\mathbf{y_1}} - \hat{\mathbf{y_2}}) = -(\hat{\mathbf{z_1}} - \hat{\mathbf{z_2}})$$

Multiplying both sides by $\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2$ and using the fact that $\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2 \perp \hat{\mathbf{z}}_1 - \hat{\mathbf{z}}_2$, we can show that $||\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2||$ and therefore the result.

From these concepts, we can easily get the following lemma:

Let $W = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ be an orthogonal basis for W. If $\mathbf{y} \in W$, then $\text{proj}_W \mathbf{y} = \mathbf{y}$.

Proposition 6.7. The Best Approximation Theorem
Let W be a subspace of $\mathbb{R}^n$. Let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, that is, for any $\mathbf{w} \in W$,

$$||\mathbf{y} - \hat{\mathbf{y}}|| \le ||\mathbf{y} - \mathbf{w}||$$

Remark: The vector $\hat{\mathbf{y}}$ is called **the best approximation** to $\mathbf{y}$ from W.

Proof. We write $\mathbf{y} - \mathbf{w} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{w})$, and the result follows from the Pythagorean theorem. Following this proposition, we are safe to define the distance between a vector and a subspace.

Definition 6.5. The **distance** from a vector $\mathbf{y}$ to a subspace W is defined as the distance from $\mathbf{y}$ to the closest point in W, that is,

$$\text{dist}(\mathbf{y}, W) = ||\mathbf{y} - \hat{\mathbf{y}}||$$

where $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto W.

Example 6.2. Find the distance from $\mathbf{y}$ to $W = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

Solution. $3\sqrt{5}$

Lastly, from the Orthogonal Decomposition Theorem, for the case when the set is an orthonormal basis, we can easily get the following lemma:

Let W be a subspace of $\mathbb{R}^n$ spanned by an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$. If $\mathbf{y} \in \mathbb{R}^n$, then
$$\text{proj}_W \mathbf{y} = \mathbf{y} \cdot \mathbf{u}_1 \mathbf{u}_1 + \cdots + \mathbf{y} \cdot \mathbf{u}_p \mathbf{u}_p.$$
If $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_p]$, then
$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y} \text{ for all } \mathbf{y} \in \mathbb{R}^n$$

Also, we may show the following lemma to be true, which involves the linear combination:

Let W be a subspace of $\mathbb{R}^n$, and $\mathbf{x}, \mathbf{y}$ are vectors in $\mathbb{R}^n$. If $\mathbf{z} = \mathbf{x} + \mathbf{y}$, then

$$\text{proj}_W \mathbf{z} = \text{proj}_W \mathbf{x} + \text{proj}_W \mathbf{y}.$$

6.4 Gram-Schmidt Process

After having the concepts of orthogonal and orthonormal sets, the thing left is to find such sets. Although we may be provided with a basis of a subspace, it may not be orthogonal or orthonormal, which limits its applications. For example, try the following exercise:

Example 6.3. Let W be the subspace of $\mathbb{R}^3$ spanned by the vectors

$$\mathbf{v}_1 = \begin{pmatrix} 3 \\ 6 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}.$$

Find an orthogonal basis for W.

In this example, it is quite easy to "guess out" an answer, but in general, we need a systematic way to find the orthogonal basis. The Gram-Schmidt process is such a simple algorithm to find an orthogonal basis for a nonzero subspace W of $\mathbb{R}^n$ given a basis for W. Remark: There should have been some reasoning behind the Gram-Schmidt process.

Proposition 6.8. The Gram-Schmidt Process
Given a basis $\{\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_p}\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_i = \begin{cases} \mathbf{x}_1 & \text{if } i = 1, \\ \mathbf{x}_i - \sum_{j=1}^{i-1} \text{proj}_{\mathbf{v}_j} \mathbf{x}_i & \text{if } i > 1. \end{cases}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is an orthogonal basis for W, and

$$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\} = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}.$$

After getting an **orthogonal basis** through the algorithm above, we can easily construct a **orthonormal basis** by normalizing each vector in the basis.

Proposition 6.9. The QR Factorization
Let A be an $m \times n$ matrix with linearly independent columns. Then there exists an $m \times n$ matrix Q with orthonormal columns and an $n \times n$ upper triangular matrix R such that $A = QR$, and The Q's columns form an orthonormal basis for the column space of A, and R is an $n \times n$ upper triangular invertible matrix with positive diagonal entries.

The proof remains to be completed. Furthermore, if columns of A are linearly dependent and $A = QR$, then R will have zero entries on its diagonal.

6.5 Least Squares Problems

Least Squares Solutions

In real application, sometimes $A\mathbf{x} = \mathbf{b}$ has no solution, and we want to find $\hat{\mathbf{x}}$ such that $A\mathbf{x}$ is as close as possible to $\mathbf{b}$, which is called the **least squares problem**.

Definition 6.6. If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, then the **least squares solution** to the inconsistent system $A\mathbf{x} = \mathbf{b}$ is the vector $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$||A\hat{\mathbf{x}} - \mathbf{b}|| \leq ||A\mathbf{x} - \mathbf{b}||$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

> **Proposition 6.10. The Least Squares Solution**
> The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the equations
> $$A^T A\mathbf{x} = A^T \mathbf{b},$$
> which is called the **normal equation** for the system $A\mathbf{x} = \mathbf{b}$.

Proof. First we prove that such $\hat{\mathbf{x}}$ exists. Denote $\hat{\mathbf{b}} = \text{proj}_{\text{Col}A}\mathbf{b}$, then since $\hat{\mathbf{b}}$ is a combination of column vectors of A, the system $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ has a solution $\hat{\mathbf{x}}$, which satisfies

$$||\mathbf{b} - A\hat{\mathbf{x}}|| = ||\mathbf{b} - \hat{\mathbf{b}}|| \leq ||\mathbf{b} - A\mathbf{x}||.$$

Next, we prove that the solution set of $A^T A\mathbf{x} = A^T \mathbf{b}$ (1) is the same as the solution set of $A\mathbf{x} = \hat{\mathbf{b}}$ (2). According to the definition of orthogonal projection, $\mathbf{b} - \hat{\mathbf{b}} \perp \text{Col}A$, which indicates that

$$A^T(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0}.$$

By substituting $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$ into the equation above, we can get (1). Conversely, from (1) we can get $0 = A^T(A\mathbf{x} - \mathbf{b})$, which implies that $(\mathbf{b} - A\mathbf{x}) \perp \text{Col}A$, so $\mathbf{b} - A\mathbf{x} \in (\text{Col}A)^\perp$. Here by the uniqueness of orthogonal decomposition, we must have $\mathbf{b} - A\mathbf{x} = \mathbf{b} - \hat{\mathbf{b}}$, which directly leads to (2). $\qquad\square$

From the reasoning arguments above, we can easily find that for an $m \times n$ matrix A the following statements are logically equivalent:

- The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b}$ in $\mathbb{R}^m$.

- The columns of A are linearly independent.

- The matrix $A^T A$ is invertible.

Alternative ways for finding the least squares solution

(1) When the columns of A are orthogonal.
Since $\text{Col}A$ is an orthogonal subspace, by definition we can write the orthogonal projection of $\mathbf{b}$ onto $\text{Col}A$ as

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col}A}\mathbf{b} = \frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1}\mathbf{a}_1 + \cdots + \frac{\mathbf{b} \cdot \mathbf{a}_n}{\mathbf{a}_n \cdot \mathbf{a}_n}\mathbf{a}_n = A \cdot (x_1, x_2, \cdots, x_n)^T,$$

where $x_i = \frac{\mathbf{b} \cdot \mathbf{a}_i}{\mathbf{a}_i \cdot \mathbf{a}_i}$. Hence the least squares solution is $\hat{\mathbf{x}} = (x_1, x_2, \cdots, x_n)^T$.

(2) When columns of A are linearly independent.
Here the least-squares solution can often be computed more reliable through a QR factorization of A. Let $A = QR$ be a QR factorization of A. Then, for each $\mathbf{b}$ in $\mathbb{R}^m$, the least-squares solution of $A\mathbf{x} = \mathbf{b}$ is given by
$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.$$

Proof. Let $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$, then

$$A\hat{\mathbf{x}} = QR\hat{\mathbf{x}} = QRR^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b}.$$

Note that the columns of Q form an orthonormal basis for $\text{Col}A$, so $QQ^T\mathbf{b}$ is the orthogonal projection of $\mathbf{b}$ onto $\text{Col}A$ - Why?. So $\hat{\mathbf{x}}$ is the least-squares solution. The uniqueness follows from the fact that the columns of A are linearly independent.

These methods requires lots of exercise to get familiar with, and the readers are encouraged to try more exercises to get a better understanding of the least squares problem. For example, following is an example which involves different methods covered in these chapters. (Please calculate this!)

Example 6.4. Find the least square solution using QR factorization for the following system:

$$\begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 7 \\ -3 \end{pmatrix}$$

This blank is left for the rough work of the example above.

6.6 Applications

Linear Regression (Least Squares lines)

In the context of linear regression, we are given a set of data points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, and we want to find the best-fitting line $y = mx + b$ to the data, where we want to minimize

$$\sum_{i=1}^{n} (y_i - (mx_i + b))^2.$$

Given the description above, we can always write

$$y_i = mx_i + b + \epsilon_i,$$

where ϵ_i is the error in the ith measurement. Hence, we can use the matrix form to represent the problem as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} b \\ m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Now we want to minimize $||\mathbf{x}\beta - \mathbf{y}||$, where

$$\mathbf{x} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} b \\ m \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

which is equivalent to find the least squares solution of the system $\mathbf{x}\beta = \mathbf{y}$.

General Linear Model

Fit data by curves that have the form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_n f_n(x),$$

where $f_0(x), f_1(x), \cdots, f_n(x)$ are known functions of x. The goal is to find the coefficients $\beta_0, \beta_1, \cdots, \beta_n$ that minimize the sum of the squares of the errors.

> **Example 6.5.** Given the data $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ that comes from a company's total costs. The equation of the form
>
> $$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$
>
> is an appropriate model for the data. Describe the linear model that gives a best fit to the data.
> **Solution.** $\mathbf{y} = \mathbf{x}\beta + \varepsilon$, where
>
> $$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{pmatrix}.$$
>
> Now we can solve the least squares solution of $\mathbf{x}\beta = \mathbf{y}$ to get β.

Note that when we have more than one independent variable, we can use the same method to find the best-fitting plane or hyperplane, such as

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 w.$$