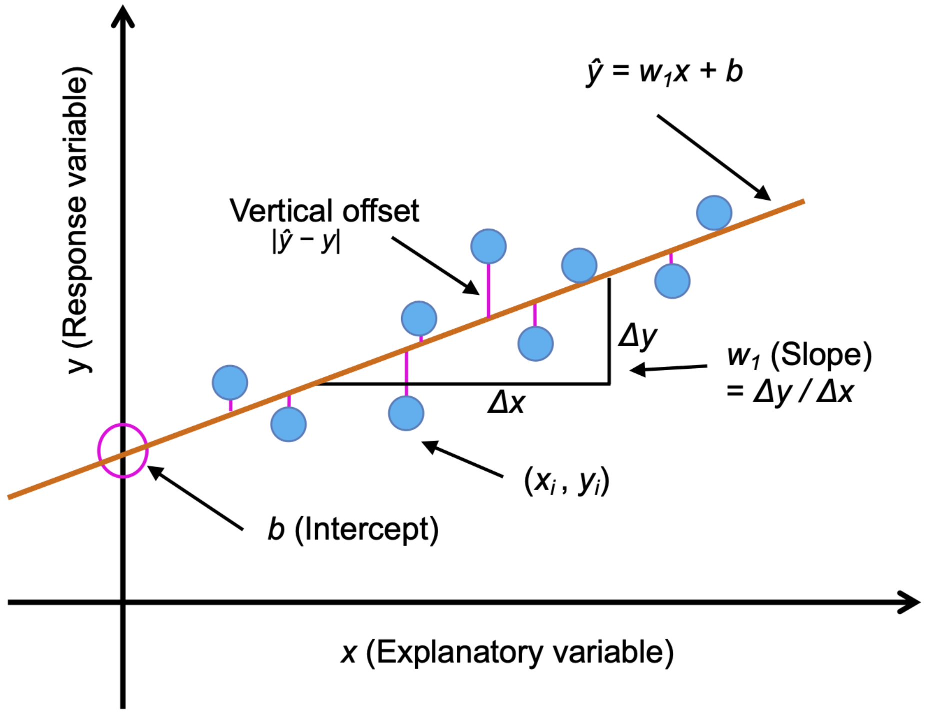
Least Squares Problems

Least Squares Problems

- The data may include errors in measurement or experiment-related inaccuracies

- We require a model to provide an optimal approximation in some sense
  - The sum of squares of errors between the y values of the data points and the corresponding y values of the approximating curve should be minimized



https://github.com/rasbt/machine-learning-book

The Ceres Orbit of Gauss

The technique of least squares was developed independently by Adrien-Marie Legendre and Carl Friedrich Gauss. The first paper on the subject was published by Legendre in 1806, although there is clear evidence that Gauss had discovered it as a student nine years prior to Legendre's paper and had used the method to do astronomical calculations. Figure 5.3.1 is a portrait of Gauss.

On January 1, 1801, the Italian astronomer Giuseppe Piazzi discovered the asteroid Ceres. He was able to track the asteroid for six weeks, but it was lost due to interference caused by the sun. A number of leading astronomers published papers predicting the orbit of the asteroid. Gauss also published a forecast, but his predicted orbit differed considerably from the others. Ceres was relocated by one observer on December 7 and by another on January 1, 1802. In both cases, the position was very close to that predicted by Gauss. Gauss won instant fame in astronomical circles and for a time was more well known as an astronomer than as a mathematician. The key to his success was the use of the method of least squares.

Carl Friedrich Gauss
(Pearson Education, Inc.)

Least Squares Solutions of Overdetermined Systems

- A least squares problem can generally be formulated as an overdetermined linear system of equations
  - More equations than unknowns
  - Such systems are usually inconsistent

- Given an $m \times n$ system $A\boldsymbol{x} = \boldsymbol{b}$ with $m > n$, we cannot expect in general to find a vector $\boldsymbol{x} \in \mathbb{R}^n$ for which $A\boldsymbol{x}$ equals $\boldsymbol{b}$
  - Instead, we can look for a vector $\boldsymbol{x}$ for which $A\boldsymbol{x}$ is closest to $\boldsymbol{b}$

- If we are given a system of equations $A\boldsymbol{x} = \boldsymbol{b}$, where A is an $m \times n$ matrix with $m > n$ and $\boldsymbol{b} \in \mathbb{R}^m$, then, for each $\boldsymbol{x} \in \mathbb{R}^n$, we can form a residual

$$r(\boldsymbol{x}) = \boldsymbol{b} - A\boldsymbol{x}$$

The distance between $\boldsymbol{b}$ and $A\boldsymbol{x}$ is given by

$$\|\boldsymbol{b} - A\boldsymbol{x}\| = \|r(\boldsymbol{x})\|$$

Least Squares Solutions of Overdetermined Systems

- We wish to find a vector $\boldsymbol{x} \in \mathbb{R}^n$ for which $\|r(\boldsymbol{x})\|$ will be a minimum
  - Minimizing $\|r(\boldsymbol{x})\|$ is equivalent to minimizing $\|r(\boldsymbol{x})\|^2$

- A vector $\hat{\boldsymbol{x}}$ that accomplishes this is said to be a *least squares solution* of the system $A\boldsymbol{x} = \boldsymbol{b}$

- If $\hat{\boldsymbol{x}}$ is a least squares solution of the system $A\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{p} = A\hat{\boldsymbol{x}}$, then $\boldsymbol{p}$ is a vector in the column space of A that is closest to $\boldsymbol{b}$

Least Squares Solutions of Overdetermined Systems

- Let S be a subspace of $\mathbb{R}^m$. For each $\boldsymbol{b} \in \mathbb{R}^m$, there is a unique element $\boldsymbol{p}$ of S that is closest to $\boldsymbol{b}$; that is

$$\|\boldsymbol{b}-\boldsymbol{y}\| > \|\boldsymbol{b}-\boldsymbol{p}\|$$

for any $\boldsymbol{y} \neq \boldsymbol{p}$ in S

- A given vector $\boldsymbol{p}$ in S will be closest to a given vector $\boldsymbol{b} \in \mathbb{R}^m$ if and only if $\boldsymbol{b}-\boldsymbol{p} \in S^{\perp}$

- Proof

Since $\mathbb{R}^m = S \oplus S^{\perp}$, each element $\mathbf{b}$ in $\mathbb{R}^m$ can be expressed uniquely as a sum

$$\mathbf{b} = \mathbf{p} + \mathbf{z}$$

where $\mathbf{p} \in S$ and $\mathbf{z} \in S^{\perp}$. If $\mathbf{y}$ is any other element of S, then

$$\|\mathbf{b}-\mathbf{y}\|^2 = \|(\mathbf{b}-\mathbf{p}) + (\mathbf{p}-\mathbf{y})\|^2$$

Since $\mathbf{p}-\mathbf{y} \in S$ and $\mathbf{b}-\mathbf{p} = \mathbf{z} \in S^{\perp}$, it follows from the Pythagorean law that

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

Therefore,

$$\|\mathbf{b}-\mathbf{y}\| > \|\mathbf{b}-\mathbf{p}\|$$

Thus, if $\mathbf{p} \in S$ and $\mathbf{b}-\mathbf{p} \in S^{\perp}$, then $\mathbf{p}$ is the element of S that is closest to $\mathbf{b}$. Conversely, if $\mathbf{q} \in S$ and $\mathbf{b}-\mathbf{q} \notin S^{\perp}$, then $\mathbf{q} \neq \mathbf{p}$, and it follows from the preceding argument (with $\mathbf{y} = \mathbf{q}$) that

$$\|\mathbf{b}-\mathbf{q}\| > \|\mathbf{b}-\mathbf{p}\| \qquad \blacksquare$$

Least Squares Solutions of Overdetermined Systems

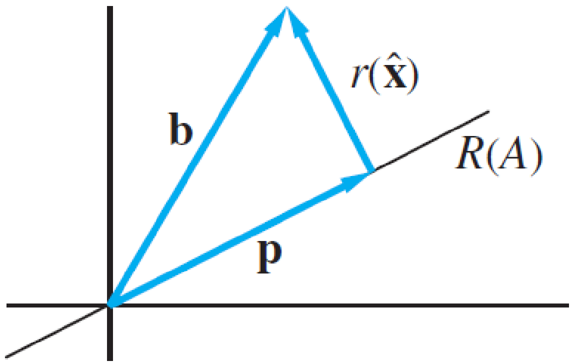
- A vector $\hat{\boldsymbol{x}}$ will be a solution of the least squares problem $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\boldsymbol{p} = A\hat{\boldsymbol{x}}$ is the vector in $R(A)$ that is closest to $\boldsymbol{b}$

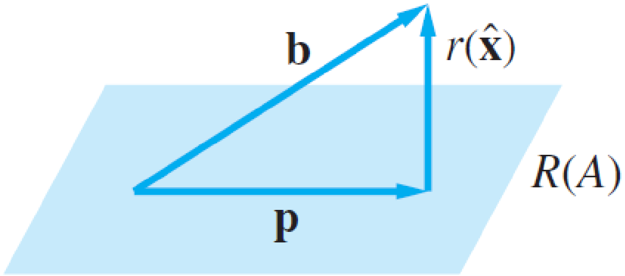
- The vector $\boldsymbol{p}$ is said to be the *projection of* $\mathbf{b}$ *onto* $R(A)$

$$\boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - A\hat{\boldsymbol{x}} = r(\hat{\boldsymbol{x}})$$

must be an element of $R(A)^{\perp}$. Thus, $\hat{\boldsymbol{x}}$ is a solution of the least squares problem if and only if

$$r(\hat{\boldsymbol{x}}) \in R(A)^{\perp}$$





(a) $\mathbf{b} \in R^2$ and A is a 2×1 matrix of rank 1.   (b) $\mathbf{b} \in R^2$ and A is a 3×2 matrix of rank 2.

Normal Equations

- A vector $\hat{\boldsymbol{x}}$ will be a least squares solution to the system $A\boldsymbol{x} = \boldsymbol{b}$ if and only if

$$r(\hat{\boldsymbol{x}}) \in N(A^T)$$

or, equivalently,

$$A^T r(\hat{\boldsymbol{x}}) = A^T(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = \mathbf{0}$$

Thus, to solve the least squares problem $A\boldsymbol{x} = \boldsymbol{b}$, we must solve

$$A^T\boldsymbol{b} = A^T A\boldsymbol{x}$$

- These equations are called the *normal equations*

- In general, it is possible to have more than one solution of the normal equations; however, if $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are both solutions, then, since the projection $\boldsymbol{p}$ of $\boldsymbol{b}$ onto $R(A)$ is unique,

$$A\hat{\boldsymbol{x}} = A\hat{\boldsymbol{y}} = \boldsymbol{p}$$

Normal Equations

- If A is an $m \times n$ matrix of rank n, the normal equations

$$A^T A\boldsymbol{x} = A^T \boldsymbol{b}$$

  have a unique solution

$$\widehat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$$

  and $\widehat{\boldsymbol{x}}$ is the unique least squares solution of the system $A\boldsymbol{x} = \boldsymbol{b}$

- Proof

We will first show that A^TA is nonsingular. To prove this, let $\mathbf{z}$ be a solution of

$$A^TA\mathbf{x} = \mathbf{0} \tag{3}$$

Then $A\mathbf{z} \in N(A^T)$. Clearly, $A\mathbf{z} \in R(A) = N(A^T)^\perp$. Since $N(A^T) \cap N(A^T)^\perp = \{\mathbf{0}\}$, it follows that $A\mathbf{z} = \mathbf{0}$. If A has rank n, the column vectors of A are linearly independent and, consequently, $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Thus, $\mathbf{z} = \mathbf{0}$ and (3) has only the trivial solution. Therefore, by Theorem 1.5.2, A^TA is nonsingular. It follows that $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$ is the unique solution of the normal equations and, consequently, the unique least squares solution of the system $A\mathbf{x} = \mathbf{b}$. ■

Normal Equations

- The projection vector

$$\boldsymbol{p} = A\hat{\boldsymbol{x}} = A(A^TA)^{-1}A^T\boldsymbol{b}$$

is the element of $R(A)$ that is closest to $\boldsymbol{b}$ in the least squares sense

- The matrix $P = A(A^TA)^{-1}A^T$ is called the *projection matrix*

Normal Equations

Ex 1. Find the least squares solution of the system

$$\begin{aligned} x_1 + x_2 &= 3 \\ -2x_1 + 3x_2 &= 1 \\ 2x_1 - x_2 &= 2 \end{aligned}$$

- Solution

The normal equations for this system are

$$\begin{pmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

This simplifies to the 2×2 system

$$\begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

The solution of the 2×2 system is $\left(\frac{83}{50}, \frac{71}{50}\right)^T$. ■

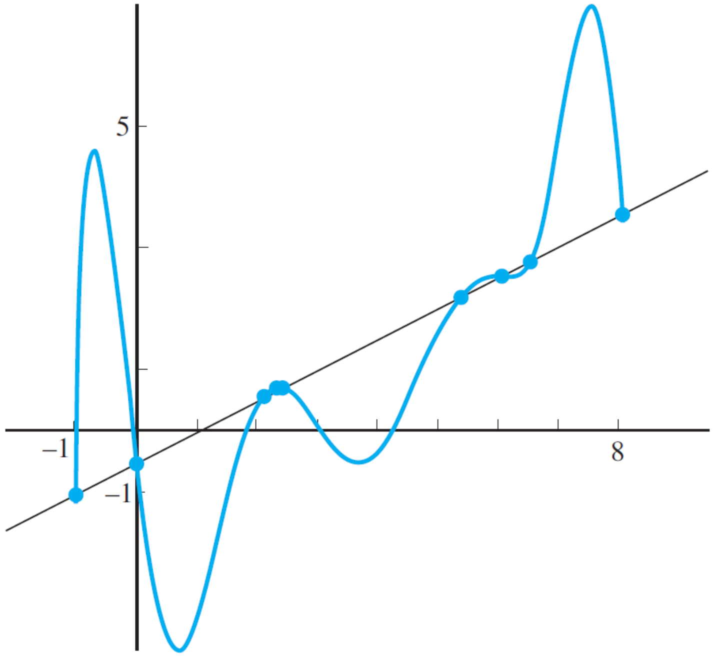
Interpolating Polynomial

- If the data consist of $n+1$ points in the plane, it is possible to find a polynomial of degree n or less passing through all the points

- Such a polynomial is called an *interpolating polynomial*

- There is no reason to require that the function pass through all the points since the data usually involve experimental error

x	–1.00	0.00	2.10	2.30	2.40	5.30	6.00	6.50	8.00
y	–1.02	–0.52	0.55	0.70	0.70	2.13	2.52	2.82	3.54



Interpolating Polynomial

Ex 2. Given the below data, find the best least squares fit by a linear function

x	0	3	6
y	1	4	5

- Solution

For this example, the system (4) becomes

$$A\mathbf{c} = \mathbf{y} \qquad\qquad y = c_0 + c_1 x$$

where

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}$$

The normal equations

$$A^T A\mathbf{c} = A^T \mathbf{y}$$

simplify to

$$\begin{bmatrix} 3 & 9 \\ 9 & 45 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 42 \end{bmatrix} \tag{5}$$

The solution of this system is $\left(\frac{4}{3}, \frac{2}{3}\right)$. Thus, the best linear least squares fit is given by

$$y = \frac{4}{3} + \frac{2}{3}x$$

Interpolating Polynomial

- Example 2 could also have been solved using calculus

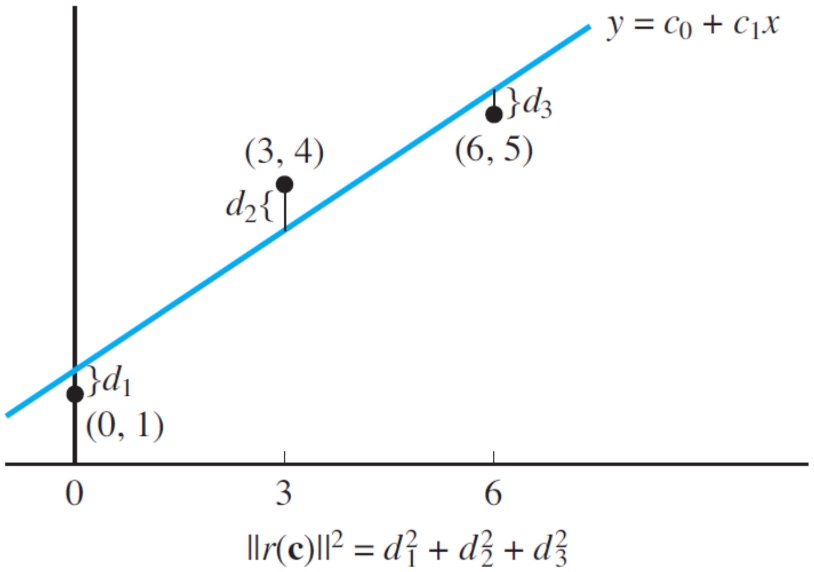
- The residual $r(\boldsymbol{c})$ is given by $r(\boldsymbol{c}) = y - A\boldsymbol{c}$

$$\begin{aligned} &\|r(\boldsymbol{c})\|^2 \\ &= \|\boldsymbol{y} - A\boldsymbol{c}\|^2 \\ &= [1-(c_0+0c_1)]^2 + [4-(c_0+3c_1)]^2 + [5-(c_0+6c_1)]^2 \\ &= f(c_0, c_1) \end{aligned}$$

- The minimum of this function will occur when its partial derivatives are zero:

$$\frac{\partial f}{\partial c_0} = -2(10 - 3c_0 - 9c_1) = 0$$
$$\frac{\partial f}{\partial c_1} = -6(14 - 3c_0 - 15c_1) = 0$$

cf. $\begin{bmatrix} 3 & 9 \\ 9 & 45 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 42 \end{bmatrix}$



Interpolating Polynomial

- If the data do not resemble a linear function, we could use a higher degree polynomial

- To find the coefficients $c_0, c_1, \cdots, c_n$ of the best least squares fit to the data

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

by a polynomial of degree n, we must find the least squares solution of the system:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Interpolating Polynomial

Ex 3. Find the best quadratic least squares fit to the data

x	0	1	2	3
y	3	2	4	4

- Solution

For this example, the system (6) becomes

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 4 \\ 4 \end{bmatrix}$$

Thus, the normal equations are

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 4 \\ 4 \end{bmatrix}$$

These simplify to

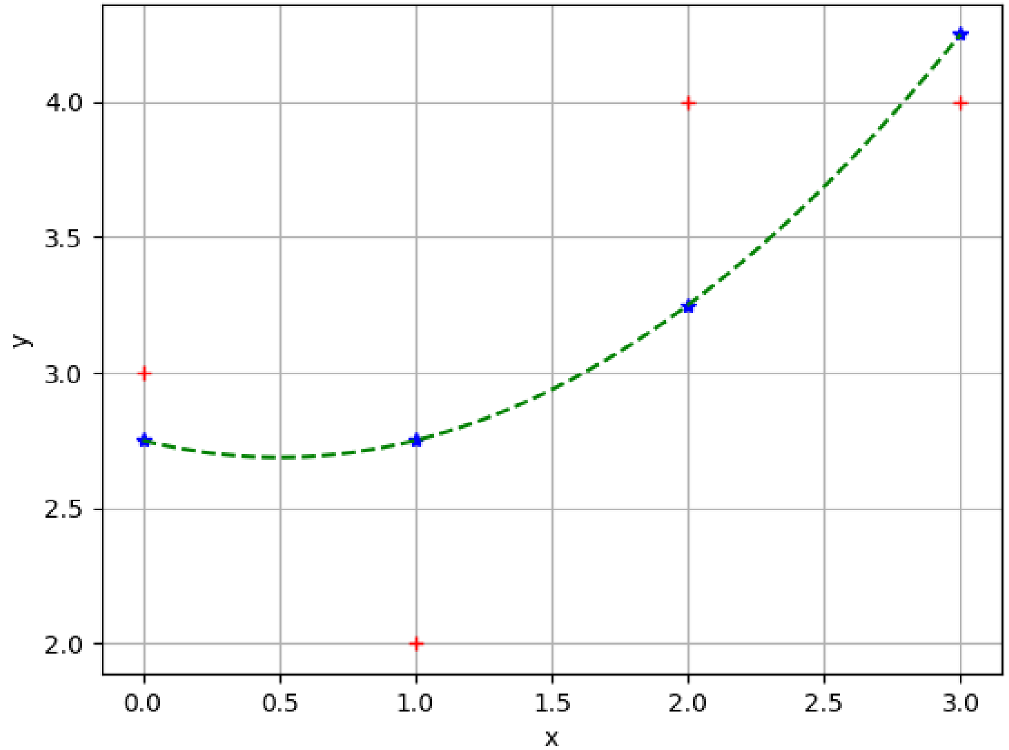
$$\begin{bmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 13 \\ 22 \\ 54 \end{bmatrix}$$

The solution of this system is $(2.75, -0.25, 0.25)$. The quadratic polynomial that gives the best least squares fit to the data is

$$p(x) = 2.75 - 0.25x + 0.25x^2$$

Interpolating Polynomial

```python
import numpy as np
from matplotlib import pyplot as plt

x = np.array([0, 1, 2, 3])
y = np.array([3, 2, 4, 4])

A = np.vstack([np.ones_like(x), x, x**2]).T

print('A:\n', A)

coef = np.linalg.lstsq(A, y, rcond=None)
y_hat = A @ coef[0]

xx = np.linspace(x[0], x[-1], 100)
X = np.vstack([np.ones_like(xx), xx, xx**2]).T
yy = X @ coef[0]

print('Least squares solution: y={:.2f}{:+.2f}x{:+.2f}x^2'.format(*coef[0]))

plt.plot(x, y, 'r+')
plt.plot(x, y_hat, 'b*')
plt.plot(xx, yy, 'g--')
plt.grid()
plt.xlabel('x')
plt.ylabel('y')
plt.show()
```




```
A:
 [[1 0 0]
 [1 1 1]
 [1 2 4]
 [1 3 9]]
Least squares solution: y=2.75-0.25x+0.25x^2
```

Exercises

12. Show that if

$$\begin{bmatrix} A & I \\ O & A^T \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}} \\ \boldsymbol{r} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix}$$

then $\hat{\boldsymbol{x}}$ is a least squares solution of the system $A\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{r}$ is the residual vector

Inner Product Spaces

Inner Product Spaces

- An *inner product* on a vector space V is an operation on V that assigns, to each pair of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in V, a real number $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ satisfying the following conditions:

  I. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ with equality if and only if $\boldsymbol{x} = \mathbf{0}$

  II. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in V

  III. $\langle \alpha \boldsymbol{x} + \beta \boldsymbol{y}, \boldsymbol{z} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \beta \langle \boldsymbol{y}, \boldsymbol{z} \rangle$

- A vector space V with an inner product is called an *inner product space*

Inner Product Spaces

- The vector space $\mathbb{R}^n$
  - The standard inner product for $\mathbb{R}^n$ is the scalar product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y}$$

  - Given a vector $\boldsymbol{w}$ with positive entries, an inner product on $\mathbb{R}^n$ is defined as follows:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i w_i$$

  - The entries w_i are referred to as *weights*

- The vector space $\mathbb{R}^{m \times n}$
  - Given A and B in $\mathbb{R}^{m \times n}$, we can define an inner product by

$$\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

Inner Product Spaces

- The vector space $C[a,b]$
  - An inner product on $C[a,b]$ is defined by

$$\langle f,g\rangle = \int_a^b f(x)g(x)dx$$

$$\langle f,f\rangle = \int_a^b \left(f(x)\right)^2 dx \geq 0$$

- The vector space P_n
  - Let $x_1, x_2, \cdots, x_n$ be distinct real numbers. For each pair of polynomials in P_n, an inner product is defined as follows:

$$\langle p,q\rangle = \sum_{i=1}^{n} p(x_i)q(x_i)$$

Basic Properties of Inner Product Spaces

- If $\boldsymbol{v}$ is a vector in an inner product space V, the *length*, or *norm* of $\boldsymbol{v}$ is given by

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

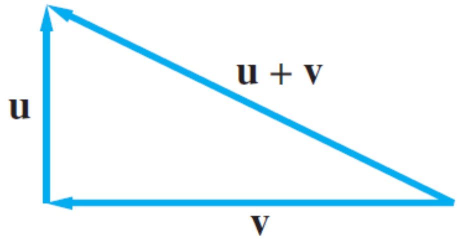
- Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be orthogonal if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$

- The Pythagorean Law
  - If $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal vectors in an inner product space V, then

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$$

  - Proof

$$\begin{aligned}\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\end{aligned}$$



Basic Properties of Inner Product Spaces

Ex 1. Consider the vector space $C[-1,1]$ with an inner product defined as follows:

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx$$

The vectors(*i.e.*, functions) 1 and x are orthogonal

- Proof

$$\langle 1, x\rangle = \int_{-1}^{1} 1 \cdot x\, dx = 0$$

To determine the lengths of these vectors, we compute

$$\langle 1, 1\rangle = \int_{-1}^{1} 1 \cdot 1\, dx = 2$$

$$\langle x, x\rangle = \int_{-1}^{1} x^2\, dx = \frac{2}{3}$$

It follows that

$$\|1\| = (\langle 1, 1\rangle)^{1/2} = \sqrt{2}$$

$$\|x\| = (\langle x, x\rangle)^{1/2} = \frac{\sqrt{6}}{3}$$

Since 1 and x are orthogonal, they satisfy the Pythagorean law:

$$\|1 + x\|^2 = \|1\|^2 + \|x\|^2 = 2 + \frac{2}{3} = \frac{8}{3}$$

The reader may verify that

$$\|1 + x\|^2 = \langle 1 + x, 1 + x\rangle = \int_{-1}^{1} (1 + x)^2\, dx = \frac{8}{3}$$

Basic Properties of Inner Product Spaces

Ex 2. For the vector space $C[-\pi,\pi]$, if we use a constant weight function $w(x)=\frac{1}{\pi}$ to define an inner product

$$\langle f,g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)dx$$

then

$$\langle \cos x, \sin x\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos x \sin x \, dx = 0$$

$$\langle \cos x, \cos x\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos x \cos x \, dx = 1$$

$$\langle \sin x, \sin x\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin x \sin x \, dx = 1$$

Thus, $\cos x$ and $\sin x$ are orthogonal unit vectors with respect to this inner product. It follows from the Pythagorean law that

$$\|\cos x + \sin x\| = \sqrt{2}$$

Power-reduction formulae [edit]

Obtained by solving the second and third versions of the cosine double-angle formula.

Sine	Cosine
$\sin^2\theta = \frac{1-\cos(2\theta)}{2}$	$\cos^2\theta = \frac{1+\cos(2\theta)}{2}$

https://en.wikipedia.org/wiki/List_of_trigonometric_identities

Basic Properties of Inner Product Spaces

- For the vector space $\mathbb{R}^{m\times n}$, the following value derived from the inner product is called the *Frobenius norm* and is denoted by $\|\cdot\|_F$

- If $A \in \mathbb{R}^{m\times n}$, then

$$\|A\|_F = (\langle A, A\rangle)^{\frac{1}{2}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{\frac{1}{2}}$$

Ex 3. If $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 3 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 1 \\ 3 & 0 \\ -3 & 4 \end{bmatrix}$, then A and B are not orthogonal

- Proof

$$\langle A, B\rangle = 1 \cdot -1 + 1 \cdot 1 + 1 \cdot 3 + 2 \cdot 0 + 3 \cdot -3 + 3 \cdot 4 = 6$$

Hence, A is not orthogonal to B. The norms of these matrices are given by

$$\|A\|_F = (1 + 1 + 1 + 4 + 9 + 9)^{1/2} = 5$$
$$\|B\|_F = (1 + 1 + 9 + 0 + 9 + 16)^{1/2} = 6$$

Basic Properties of Inner Product Spaces

- If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in an inner product space V and $\boldsymbol{v} \neq \boldsymbol{0}$, then the *scalar projection* of $\boldsymbol{u}$ onto $\boldsymbol{v}$ is given by

$$\alpha = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|}$$

and the *vector projection* of $\boldsymbol{u}$ onto $\boldsymbol{v}$ is given by

$$\boldsymbol{p} = \alpha \left(\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right) = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \boldsymbol{v}$$

Basic Properties of Inner Product Spaces

- Observations
  - If $\boldsymbol{v} \neq \mathbf{0}$ and $\boldsymbol{p}$ is the vector projection of $\boldsymbol{u}$ onto $\boldsymbol{v}$, then

    I. $\boldsymbol{u} - \boldsymbol{p}$ and $\boldsymbol{p}$ are orthogonal

    II. $\boldsymbol{u} = \boldsymbol{p}$ if and only if $\boldsymbol{u}$ is a scalar multiple of $\boldsymbol{v}$

- Proof

Proof of Observation I Since

$$\langle \mathbf{p}, \mathbf{p} \rangle = \langle \frac{\alpha}{\|\mathbf{v}\|}\mathbf{v}, \frac{\alpha}{\|\mathbf{v}\|}\mathbf{v} \rangle = \left(\frac{\alpha}{\|\mathbf{v}\|}\right)^2 \langle \mathbf{v}, \mathbf{v} \rangle = \alpha^2$$

and

$$\langle \mathbf{u}, \mathbf{p} \rangle = \frac{(\langle \mathbf{u}, \mathbf{v} \rangle)^2}{\langle \mathbf{v}, \mathbf{v} \rangle} = \alpha^2$$

it follows that

$$\langle \mathbf{u} - \mathbf{p}, \mathbf{p} \rangle = \langle \mathbf{u}, \mathbf{p} \rangle - \langle \mathbf{p}, \mathbf{p} \rangle = \alpha^2 - \alpha^2 = 0$$

Therefore, $\mathbf{u} - \mathbf{p}$ and $\mathbf{p}$ are orthogonal.

Proof of Observation II If $\mathbf{u} = \beta\mathbf{v}$, then the vector projection of $\mathbf{u}$ onto $\mathbf{v}$ is given by

$$\mathbf{p} = \frac{\langle \beta\mathbf{v}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v} = \beta\mathbf{v} = \mathbf{u}$$

Conversely, if $\mathbf{u} = \mathbf{p}$, it follows from (7) that

$$\mathbf{u} = \beta\mathbf{v} \quad \text{where} \quad \beta = \frac{\alpha}{\|\mathbf{v}\|}$$ ■

The Cauchy-Schwarz Inequality

- If $\boldsymbol{u}$ and $\boldsymbol{v}$ are any two vectors in an inner product space V, then

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|$$

Equality holds if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent

- Proof

If $\mathbf{v} = \mathbf{0}$, then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| = 0 = \|\mathbf{u}\| \|\mathbf{v}\|$$

If $\mathbf{v} \neq \mathbf{0}$, then let $\mathbf{p}$ be the vector projection of $\mathbf{u}$ onto $\mathbf{v}$. Since $\mathbf{p}$ is orthogonal to $\mathbf{u} - \mathbf{p}$, it follows from the Pythagorean law that

$$\|\mathbf{p}\|^2 + \|\mathbf{u} - \mathbf{p}\|^2 = \|\mathbf{u}\|^2$$

Thus,

$$\frac{(\langle \mathbf{u}, \mathbf{v} \rangle)^2}{\|\mathbf{v}\|^2} = \|\mathbf{p}\|^2 = \|\mathbf{u}\|^2 - \|\mathbf{u} - \mathbf{p}\|^2$$

and hence

$$(\langle \mathbf{u}, \mathbf{v} \rangle)^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{p}\|^2 \|\mathbf{v}\|^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \qquad (9)$$

Therefore,

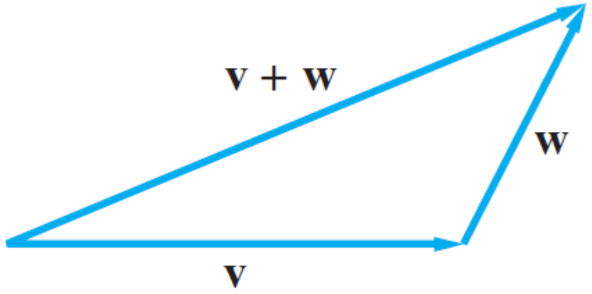
$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

Equality holds in (9) if and only if $\mathbf{u} = \mathbf{p}$. It follows from observation II that equality will hold in (8) if and only if $\mathbf{v} = \mathbf{0}$ or $\mathbf{u}$ is a multiple of $\mathbf{v}$. More simply stated, equality will hold if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. ∎

Norms

- A vector space V is said to be a *normed linear space* if, to each vector $\boldsymbol{v} \in V$, there is associated a real number $\|\boldsymbol{v}\|$, called the *norm* of $\boldsymbol{v}$, satisfying

  I. $\|\boldsymbol{v}\| \geq 0$ with equality if and only if $\boldsymbol{v} = 0$
  II. $\|\alpha\boldsymbol{v}\| = |\alpha|\|\boldsymbol{v}\|$ for any scalar α
  III. $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$ (triangle inequality)



- If V is an inner product space, then then equation $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}$ defines a norm on V

- Proof

It is easily seen that conditions **I** and **II** of the definition are satisfied. We leave this for the reader to verify and proceed to show that condition **III** is satisfied.

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \quad \text{(Cauchy–Schwarz)} \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2\end{aligned}$$

Thus,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \qquad \blacksquare$$

Norms

- L_1-norm

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

for every $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$

- Infinity norm (or uniform norm)

$$\|\boldsymbol{x}\|_\infty = \max_{1 \le i \le n} |x_i|$$

- L_p-norm

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

for any real number $p \ge 1$

Norms

- In the case of a norm that is not derived from an inner product, the Pythagorean law will not hold

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \boldsymbol{x}_2 = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

- $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal; however

$$\|\boldsymbol{x}_1\|_\infty^2 + \|\boldsymbol{x}_2\|_\infty^2 = 4 + 16 = 20 \neq \|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_\infty^2 = 16$$

- If, however, $\|\cdot\|_2$ is used, then

$$\|\boldsymbol{x}_1\|_2^2 + \|\boldsymbol{x}_2\|_2^2 = 5 + 20 = 25 = \|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_2^2$$

Ex 5. Let $\boldsymbol{x}$ be the vector $(4, -5, 3)^T \in \mathbb{R}^3$. Compute $\|\boldsymbol{x}\|_1$, $\|\boldsymbol{x}\|_2$, and $\|\boldsymbol{x}\|_\infty$

$$\|\mathbf{x}\|_1 = |4| + |-5| + |3| = 12$$

$$\|\mathbf{x}\|_2 = \sqrt{16 + 25 + 9} = 5\sqrt{2}$$

$$\|\mathbf{x}\|_\infty = \max(|4|,\ |-5|,\ |3|) = 5$$

Norms

- Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors in a normed space. The *distance* between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined to be the number $\|\boldsymbol{y}-\boldsymbol{x}\|$