

Because $\vec{a}_i^T \vec{a}_i = 1$, $\beta_i = 0$ for the equation to hold. Then, since this is true for all i from 1 to N , all the elements of the vector beta must be zero ($\vec{\beta} = \vec{0}$). Because $\vec{x} = \vec{0}$ implies $\vec{\beta} = \vec{0}$, the columns of \mathbf{A} are linearly independent.

Now, we will show that any vector $\vec{x} \in \mathbb{R}^N$ can be represented as a linear combination of the columns of \mathbf{A} .

$$\vec{x} = \mathbf{A}\vec{\beta} = \beta_1 \vec{a}_1 + \dots + \beta_N \vec{a}_N \quad (7)$$

Because we know that the N columns of \mathbf{A} are linearly independent, then there exists \mathbf{A}^{-1} . Applying the inverse to the equation above,

$$\mathbf{A}^{-1} \mathbf{A} \vec{\beta} = \mathbf{A}^{-1} \vec{x} \quad (8)$$

$$\vec{\beta} = \mathbf{A}^{-1} \vec{x}, \quad (9)$$

we find that there exists a unique β that allow us to represent any \vec{x} as a linear combination of the columns of \mathbf{A} .

- (c) When $\mathbf{A} \in \mathbb{R}^{N \times M}$ and $N \geq M$ (i.e. tall matrices), show that if the matrix is orthonormal, then $\mathbf{A}^T \mathbf{A} = \mathbf{I}_{M \times M}$.

Answer: Want to show $\mathbf{A}^T \mathbf{A} = \mathbf{I}_{M \times M}$.

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} \vec{a}_1^T \vec{a}_1 & \vec{a}_1^T \vec{a}_2 & \dots & \vec{a}_1^T \vec{a}_n \\ \vec{a}_2^T \vec{a}_1 & \vec{a}_2^T \vec{a}_2 & \dots & \vec{a}_2^T \vec{a}_n \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} = \mathbf{I}_{M \times M} \quad (10)$$

When $\vec{a}_i^T \vec{a}_i = \|\vec{a}_i\|^2 = 1$ and when $i \neq j$, $\vec{a}_i^T \vec{a}_j = 0$ because the column vectors are orthogonal.

- (d) Again, suppose $\mathbf{A} \in \mathbb{R}^{N \times M}$ where $N \geq M$ is an orthonormal matrix. Show that the projection of \vec{y} onto the subspace spanned by the columns of \mathbf{A} is now $\mathbf{A} \mathbf{A}^T \vec{y}$.

Answer:

Starting with the result from part (a),

$$\mathbf{A} \hat{\vec{x}} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \vec{y}, \quad (11)$$

we can apply the result from part (c),

$$\mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \vec{y} = \mathbf{A} \mathbf{I} \mathbf{A}^T \vec{y} \quad (12)$$

$$= \mathbf{A} \mathbf{A}^T \vec{y} \quad (13)$$

- (e) Given $\mathbf{A} \in \mathbb{R}^{N \times M} = \begin{bmatrix} 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and the columns of \mathbf{A} are orthonormal, find the least squares solution to $\mathbf{A} \hat{\vec{x}} = \vec{y}$ where $\vec{y} = [5 \ 12 \ 7 \ 8]^T$.

Answer:

Method 1:

Since the columns of \mathbf{A} are orthonormal, from part (d) we know that

$$\hat{\mathbf{x}} = \mathbf{A}^T \vec{y} = \begin{bmatrix} \langle \vec{a}_1, \vec{y} \rangle \\ \langle \vec{a}_2, \vec{y} \rangle \\ \langle \vec{a}_3, \vec{y} \rangle \end{bmatrix}.$$

Note that this is equivalent to projecting \vec{y} onto each column of \mathbf{A} :

$$\hat{x}_1 = \frac{\langle \vec{a}_1, \vec{y} \rangle}{\|\vec{a}_1\|^2} = \langle \vec{a}_1, \vec{y} \rangle = 8$$

$$\hat{x}_2 = \frac{\langle \vec{a}_2, \vec{y} \rangle}{\|\vec{a}_2\|^2} = \langle \vec{a}_2, \vec{y} \rangle = 7$$

$$\hat{x}_3 = \frac{\langle \vec{a}_3, \vec{y} \rangle}{\|\vec{a}_3\|^2} = \langle \vec{a}_3, \vec{y} \rangle = \frac{17\sqrt{2}}{2}$$

Method 2 (Alternatively you can use the least squares formula):

$$\begin{aligned} \hat{\mathbf{x}} &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \vec{y} = \left(\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \\ 7 \\ 8 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \\ 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ \frac{17\sqrt{2}}{2} \end{bmatrix} \end{aligned}$$

2. Polynomial Fitting

Let's try an example. Say we know that the output, y , is a quartic polynomial in x . This means that we know that y and x are related as follows:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$$

We're also given the following observations:

x	y
0.0	24.0
0.5	6.61
1.0	0.0
1.5	-0.95
2.0	0.07
2.5	0.73
3.0	-0.12
3.5	-0.83
4.0	-0.04
4.5	6.42

(a) What are the unknowns in this question? What are we trying to solve for?

Answer:

The unknowns are a_0 , a_1 , a_2 , a_3 , and a_4 . They are also what we are trying to solve for.

- (b) Can you write an equation corresponding to the first observation (x_0, y_0) , in terms of a_0, a_1, a_2, a_3 , and a_4 ? What does this equation look like? Is it linear in the unknowns?

Answer:

Plugging (x_0, y_0) into the expression for y in terms of x , we get

$$24 = a_0 + a_1 \cdot 0 + a_2 \cdot 0^2 + a_3 \cdot 0^3 + a_4 \cdot 0^4$$

You can see that this equation is linear in a_0, a_1, a_2, a_3 , and a_4 .

- (c) Now, write a system of equations in terms of a_0, a_1, a_2, a_3 , and a_4 using *all of the observations*.

Answer:

Write the next equation using the second observation. You will now get:

$$6.61 = a_0 + a_1 \cdot (0.5) + a_2 \cdot (0.5)^2 + a_3 \cdot (0.5)^3 + a_4 \cdot (0.5)^4$$

And for the third:

$$0.0 = a_0 + a_1 \cdot (1) + a_2 \cdot 1^2 + a_3 \cdot 1^3 + a_4 \cdot 1^4$$

Do you see a pattern? Let's write the entire system of equations in terms of a matrix now.

$$\begin{bmatrix} 1 & 0 & 0^2 & 0^3 & 0^4 \\ 1 & 0.5 & (0.5)^2 & (0.5)^3 & (0.5)^4 \\ 1 & 1 & 1^2 & 1^3 & 1^4 \\ 1 & 1.5 & (1.5)^2 & (1.5)^3 & (1.5)^4 \\ 1 & 2 & 2^2 & 2^3 & 2^4 \\ 1 & 2.5 & (2.5)^2 & (2.5)^3 & (2.5)^4 \\ 1 & 3 & 3^2 & 3^3 & 3^4 \\ 1 & 3.5 & (3.5)^2 & (3.5)^3 & (3.5)^4 \\ 1 & 4 & 4^2 & 4^3 & 4^4 \\ 1 & 4.5 & (4.5)^2 & (4.5)^3 & (4.5)^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 24 \\ 6.61 \\ 0.0 \\ -0.95 \\ 0.07 \\ 0.73 \\ -0.12 \\ -0.83 \\ -0.04 \\ 6.42 \end{bmatrix}$$

- (d) Finally, solve for a_0, a_1, a_2, a_3 , and a_4 using IPython. You have now found the quartic polynomial that best fits the data!

Answer:

Let \mathbf{D} be the big matrix from the previous part.

$$\vec{a} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \vec{y} = \begin{bmatrix} 24.00958042 \\ -49.99515152 \\ 35.0039627 \\ -9.99561772 \\ 0.99841492 \end{bmatrix}$$

It turns out that the actual parameters for the polynomial equation were:

$$\vec{a} = \begin{bmatrix} 24 \\ -50 \\ 35 \\ -10 \\ 1 \end{bmatrix}$$

(Remember that our observations were noisy.)

Therefore, we have actually done pretty well with the least squares estimate!