Ortogonal dekomposition. Minstakvadratmetoden.

Nästa sats är en utvidgning av begreppet ortogonal projektion av en vektor på en annan vektor.

Ortogonal projektion på ett underrum. Satsen om ortogonal dekomposition av en vektor. Sats 6.3.8, sid. 366

Låt W vara ett underrum av $\mathbb{R}^n$.

Alla elementen $\mathbf{y}$ ur $\mathbb{R}^n$ kan vara framställda på ett entydigt sätt som summa

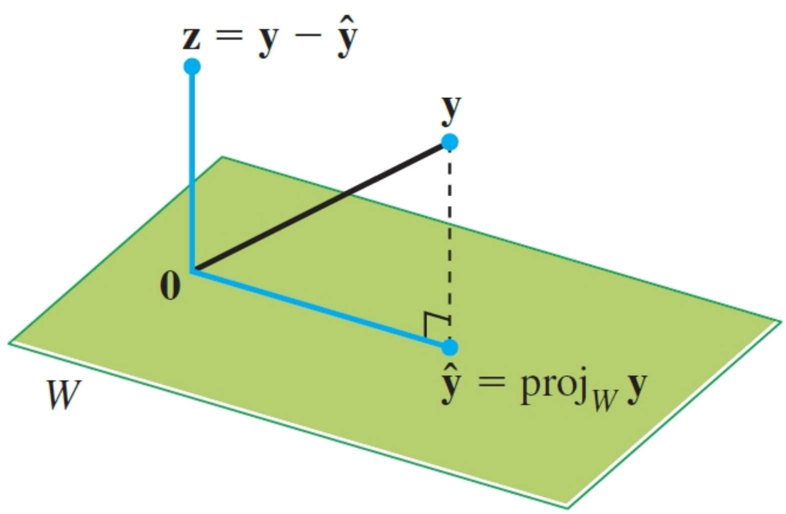
$$\begin{aligned} \mathbf{y} &= \widehat{\mathbf{y}} + \mathbf{z} = \mathrm{proj}_W \mathbf{y} + \mathbf{z} \\ \widehat{\mathbf{y}} &\in W; \qquad \mathbf{z} \in W^{\perp} \end{aligned}$$

där $\widehat{\mathbf{y}} \overset{\mathrm{def}}{=} \mathrm{proj}_W \mathbf{y}$ hör till W och $\mathbf{z}$ hör till dess ortogonala komplementet $W^{\perp}$.

Låt $\{\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_p\}$ vara en godtycklig **ortogonal bas** till W. Då kan $\widehat{\mathbf{y}}$ representeras som

$$\begin{aligned} \widehat{\mathbf{y}} &= \mathrm{proj}_W \mathbf{y} = \left(\frac{y \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \right) \mathbf{u}_1 + \left(\frac{y \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \right) \mathbf{u}_2 + ... + \left(\frac{y \cdot \mathbf{u}_p}{\|\mathbf{u}_p\|^2} \right) \mathbf{u}_p \\ &= \mathrm{proj}_{\mathbf{u}_1} \mathbf{y} + \mathrm{proj}_{\mathbf{u}_2} \mathbf{y} + ... + \mathrm{proj}_{\mathbf{u}_p} \mathbf{y} \end{aligned}$$

och $\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}$.$\square$



Ett utkast av bevis.(krävs inte på tentan)

Det är lätt att se att $\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}$ är ortogonal mot alla basvektorer $\mathbf{u}_k$, $k = 1, ..., p$ i W och måste då höra till $W^{\perp}$.

Entydigheten av framställningen $\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$ följer från följande observation. Förutsätt att det finns två olika sådanaframställningar $\mathbf{y} = \widehat{\mathbf{y}}_1 + \mathbf{z}_1$, $\mathbf{y} = \widehat{\mathbf{y}}_2 + \mathbf{z}_2$. Detta medför att $\widehat{\mathbf{y}}_1 + \mathbf{z}_1 = \widehat{\mathbf{y}}_2 + \mathbf{z}_2$ och

$$\widehat{\mathbf{y}}_1 - \widehat{\mathbf{y}}_2 = \mathbf{z}_2 - \mathbf{z}_1.$$

Vektorer i vänster hör till W och vektorer i högerm hör till $W^{\perp}$ som har bara en gemensamm punkt - noll: 0.

Detta medför att $\widehat{\mathbf{y}}_1 - \widehat{\mathbf{y}}_2 = 0$ och $\mathbf{z}_2 - \mathbf{z}_1 = 0$.

■

Lägg märke till att för $\mathbf{y} \in W$ är projektionen av $\mathbf{y}$ på W naturligtvist lika med $\mathbf{y}$: $\text{proj}_W \mathbf{y} = \mathbf{y}$.

Satsen om bästa approximationen. Sats 6.3.9, sid. 368

Låt W vara ett underrum av $\mathbb{R}^n$, $\mathbf{y}$ vara en godtycklig vektor ur $\mathbb{R}^n$ och $\widehat{\mathbf{y}} = \text{proj}_W \mathbf{y}$ vara dess ortogonala projektionen på W. Då är projektionen $\widehat{\mathbf{y}} = \text{proj}_W \mathbf{y}$ den närmaste punkt i W till punkten $\mathbf{y}$.

$$\|\mathbf{y} - \widehat{\mathbf{y}}\| \leq \|\mathbf{y} - \mathbf{v}\|$$

för alla $\mathbf{v} \in W$.

□

Bevis (krävs inte på tentan)

Välj en vektor $\mathbf{v} \in W$ inte lika med $\widehat{\mathbf{y}}$. Vektorn $\mathbf{y} - \widehat{\mathbf{y}}$ är ortogonal mot W och speciellt är den ortogonal mot $\mathbf{v} \in W$.

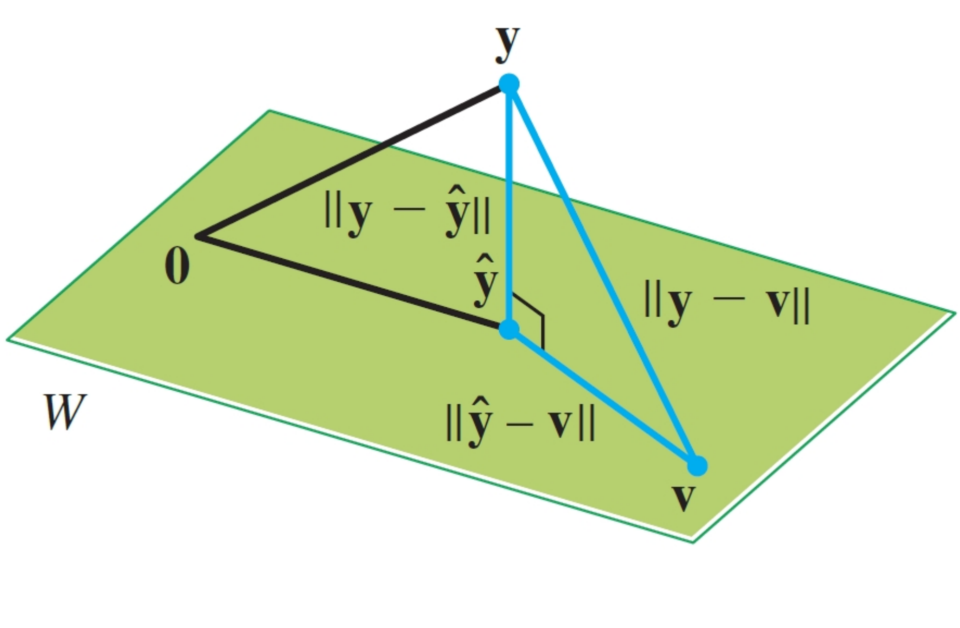
$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \widehat{\mathbf{y}}) + (\widehat{\mathbf{y}} - \mathbf{v})$$

där $(\mathbf{y} - \widehat{\mathbf{y}})$ och $(\widehat{\mathbf{y}} - \mathbf{v})$ är ortogonala mot varandra, som medför enligt Pythagor satsen att

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

Detta medför att $\|\mathbf{y} - \widehat{\mathbf{y}}\| \leq \|\mathbf{y} - \mathbf{v}\|$ eftersom $\mathbf{v} \neq \widehat{\mathbf{y}}$ och $0 < \|\widehat{\mathbf{y}} - \mathbf{v}\|$

■



Satsen om projektionen på ett underrum med hjälp av en ortonormal bas. Sats 6.3.10, sid. 369

Låt W vara ett underrum av $\mathbb{R}^n$av dimension p och $\{\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_p\}$ vara en godtycklig **orto-NORMAL bas** till W, som betyder att $\mathbf{u}_i \cdot \mathbf{u}_k = 0$, $i \neq k$, $\|\mathbf{u}_i\| = 1$, $i, = 1, ..., p$.

Då gäller två framställningar av $\text{proj}_W \mathbf{y}$ för en godtycklig vektor $\mathbf{y} \in \mathbb{R}^n$:

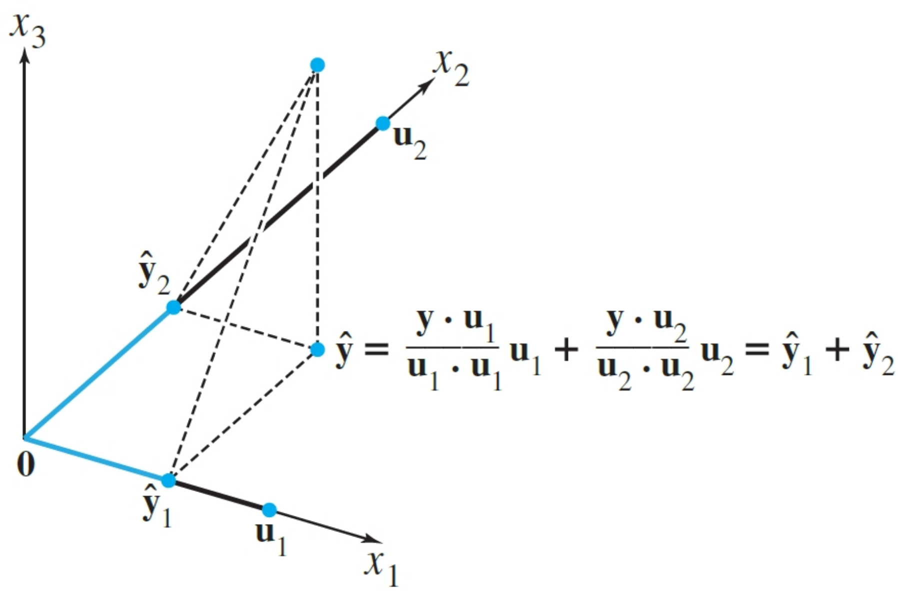
$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\, \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\, \mathbf{u}_2 + ... + (\mathbf{y} \cdot \mathbf{u}_p)\, \mathbf{u}_p$$

om $U = [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_p]\, [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_p]^T$ så kan samma uttryck framställas som matrisprodukt

$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y}, \qquad \forall \mathbf{y} \in \mathbb{R}^n$$

□

Följande bild illustrerar formeln $\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\, \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\, \mathbf{u}_2$ i fallet då W har dimensionen 2.



Gramm-Schmidt metoden för att skapa en ortogonal bas från en vanlig bas.

Ortogonala baser och ortonormala baser r lämpliga för att beräkna projektioner. Om man har en godtycklig bas till ett underrum eller till hela $\mathbb{R}^n$, så ger Gram-Schmidt metoden möjlighet att hitta en ortognal till W, och sedan även en ortonormal bas (med att normera varje vektor i ortogonala basen).

Gramm-Schmidt metoden Sats 6.4.11, sid. 373.

Låt W vara ett icke noll underrum av $\mathbb{R}^n$av dimension p och $\{\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_p\}$ vara en godtycklig bas till W. Följande beräkningsprocess ger en ny bas $\{\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_p\}$ till W som är ortogonal bas.

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \left(\frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \left(\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 \\
& \dots \\
\mathbf{v}_p &= \mathbf{x}_p - \left(\frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \dots - \left(\frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\|\mathbf{v}_{p-1}\|^2}\right)\mathbf{v}_{p-1}
\end{aligned}$$

Mera kompakt uttryck i geometriska termer med projektioner på $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{p-1}$

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \mathrm{proj}_{\mathbf{v}_1}\mathbf{x}_2 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \mathrm{proj}_{\mathbf{v}_1}\mathbf{x}_3 - \mathrm{proj}_{\mathbf{v}_2}\mathbf{x}_3 \\
& \dots \\
\mathbf{v}_p &= \mathbf{x}_p - \mathrm{proj}_{\mathbf{v}_1}\mathbf{x}_p - \dots - \mathrm{proj}_{\mathbf{v}_{p-1}}\mathbf{x}_p
\end{aligned}$$

På varje steg k beräknas komponentet $\mathbf{v}_k$ av vektorn $\mathbf{x}_k$, ortogonalt mot $W_{k-1} = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k-1}\}$ som

$$\mathbf{v}_k = \mathbf{x}_k - \mathrm{proj}_{Wk-1}\mathbf{x}_k$$

Detta gör att varje nya basvektorn $\mathbf{v}_k$ är ortogonal till tidigare byggda ortogonala basvektorer $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k-1}\}$. I slutet av beräkningsprocessen skapas p ortogonala vektorer som bygger en ortogonal bas i W.$\square$

Exempel.

EXAMPLE 2 Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

SOLUTION

Step 1. Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \mathrm{Span}\{\mathbf{x}_1\} = \mathrm{Span}\{\mathbf{v}_1\}$.

Step 2. Let $\mathbf{v}_2$ be the vector produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is, let

$$\begin{aligned}
\mathbf{v}_2 &= \mathbf{x}_2 - \operatorname{proj}_{W_1}\mathbf{x}_2 \\
&= \mathbf{x}_2 - \frac{\mathbf{x}_2\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 \qquad \text{Since } \mathbf{v}_1 = \mathbf{x}_1 \\
&= \begin{bmatrix}0\\1\\1\\1\end{bmatrix} - \frac{3}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix}
\end{aligned}$$

As in Example 1, $\mathbf{v}_2$ is the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

Step 2′ (optional). If appropriate, scale $\mathbf{v}_2$ to simplify later computations. Since $\mathbf{v}_2$ has fractional entries, it is convenient to scale it by a factor of 4 and replace $\{\mathbf{v}_1, \mathbf{v}_2\}$ by the orthogonal basis

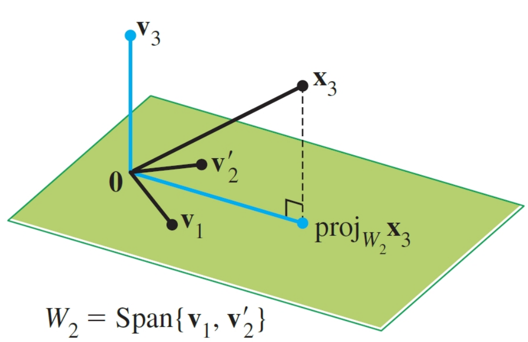
$$\mathbf{v}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad \mathbf{v}_2' = \begin{bmatrix}-3\\1\\1\\1\end{bmatrix}$$

Step 3. Let $\mathbf{v}_3$ be the vector produced by subtracting from $\mathbf{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$ to compute this projection onto W_2:

$$\operatorname{proj}_{W_2}\mathbf{x}_3 = \overset{\substack{\text{Projection of}\\ \mathbf{x}_3 \text{ onto } \mathbf{v}_1\\ \downarrow}}{\boxed{\frac{\mathbf{x}_3\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1}} + \overset{\substack{\text{Projection of}\\ \mathbf{x}_3 \text{ onto } \mathbf{v}_2'\\ \downarrow}}{\boxed{\frac{\mathbf{x}_3\cdot\mathbf{v}_2'}{\mathbf{v}_2'\cdot\mathbf{v}_2'}\mathbf{v}_2'}} = \frac{2}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \frac{2}{12}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\2/3\\2/3\\2/3\end{bmatrix}$$

Then $\mathbf{v}_3$ is the component of $\mathbf{x}_3$ orthogonal to W_2, namely,

$$\mathbf{v}_3 = \mathbf{x}_3 - \operatorname{proj}_{W_2}\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix} - \begin{bmatrix}0\\2/3\\2/3\\2/3\end{bmatrix} = \begin{bmatrix}0\\-2/3\\1/3\\1/3\end{bmatrix}$$



Minsta kvadrat problem.

Det uppstår i tillämplningar linjära ekvationer $A\mathbf{x} = \mathbf{b}$ som saknar lösningar på grund av att

systemet innehåller flera ekvationer jämfört med antalet obekanta. Speciellt uppstår sådana ekvationer när man vill approximera experimentälla data med enkla formler som innehåller ett litet antal parametrar.

Man introducerar då begreppet minstakvadratlösning som ger en approximation $\widehat{\mathbf{x}}$ av lösningar sådan att $\|A\mathbf{x} - \mathbf{b}\|$ är så liten som möjligt.

Definition.

Om A är en $m \times n$ matris och $\mathbf{b} \in \mathbb{R}^m$, en mistakvadrat lösning till systemet $A\mathbf{x} = \mathbf{b}$ är en vektor $\widehat{\mathbf{x}}$ ur $\mathbb{R}^n$ sådan att

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

för alla $\mathbf{x} \in \mathbb{R}^n$.$\square$

Minsta kvadrat problem. Satsen om normal ekvation för minsta kvadrat lösning. Sats 6.5.13.

Mängden av minstakvadratlösningar till $A\mathbf{x} = \mathbf{b}$ är samma $\iff$ som lösningsmängden till ekvationen

$$\left(A^T A\right)\widehat{\mathbf{x}} = A^T\mathbf{b}.$$

Den sista ekvationen kallas **normalekvationen till** $Ax = b$.

$\square$

Bevis $\Longrightarrow$**(krävs inte på tentan)**

Satsen om bästa approximationen påstår faktiskt att

$$\widehat{\mathbf{b}} \overset{def}{=} \operatorname{proj}_{\operatorname{Col}(A)}\mathbf{b}$$

är en punkt på $\operatorname{Col}(A)$ som ligger närmast $\mathbf{b}$.

Det betyder att ekvationen $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$ har en lösning och ger en minstakvadratlösning till ekvationen $A\mathbf{x} = \mathbf{b}$.

Satsen om ortogonal dekomposition medför att

$$\mathbf{b} - \operatorname{proj}_{\operatorname{Col}(A)}\mathbf{b} = \mathbf{b} - \widehat{\mathbf{b}} = \mathbf{b} - A\widehat{\mathbf{x}}$$

måste vara ortogonal mot $\operatorname{Col}(A)$, speciellt ortogonal mot alla kolonner i A, som kan uttryckas explicit som

$$A^T\left(\mathbf{b} - A\widehat{\mathbf{x}}\right) = 0$$

eftersom rader i A^T är samma som kolonner i A. Detta medför normalekvationen. $\blacksquare$

Implikationen åt motsatta håll genomförs med liknande argument och använder entydigheten av ortoganl dekomposition av en vektor.

Satsen om entydiga minsta kvadrat lösningar. Sats. 6.5.14, sid. 381

Låt A vara en $m \times n$ matris. Följande påståenden är ekvivalenta.

a. Ekvationen $A\mathbf{x} = \mathbf{b}$ har en entydig minstakvadratlösning för varje $\mathbf{b}$ ur $\mathbb{R}^n$

b. Kolonner i A är linjärt oberoende.

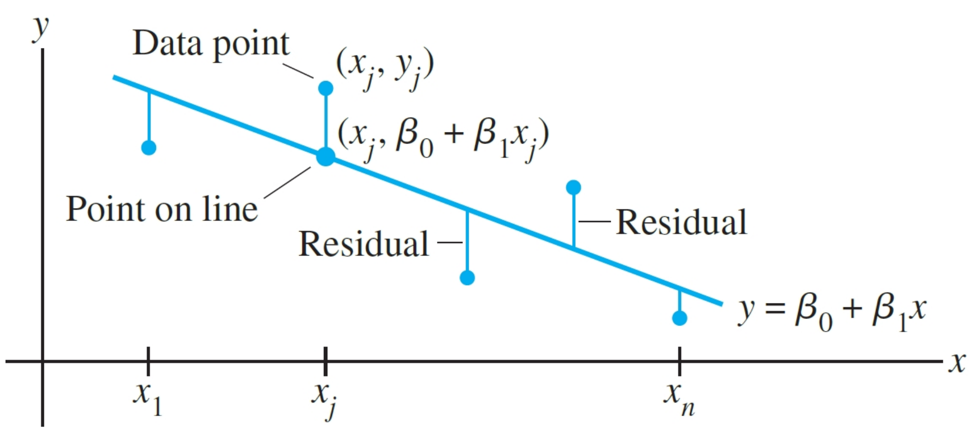
c. matrisen $A^T A$ är inverterbar.

$\square$

Bevis föreslås som övningar 6.19-21 i Lay.

Tillämpningar av minstakvadrat metoden för data anpassning. Linjära modeller. Allmänna modeller. Exempel sid. 388-391

Minstakvadrat linjer. sid. 387 i Lay



Vi har m experimentälla värden av argument x och samma antal experimentella värden av en experimentell funktion y. Vi vill approximera dem med en linjär funktion $y = \beta_0 + \beta_1 x$

Experimentella argument	Linjär approximation	Experimentälla funktions värden
x_1	$\beta_0 + \beta_1 x_1$	$= y_1$
x_2	$\beta_0 + \beta_1 x_2$	$= y_2$
x_m	$\beta_0 + \beta_1 x_m$	$= y_m$

Vi vill hitta koefficienter β_0 , β_1 som ger minimala summan av residualer (avvikelser)

$$\epsilon_k = (\beta_0 + \beta_1 x_k) - y_k$$

i kvadrat: $\sum_{k=1}^{m} \left((\beta_0 + \beta_1 x_k) - y_k\right)^2$

Det är samma som att hitta minstakvadratlösning till ekvationen

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

med matrisen och vektorer

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} ; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} ; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Matrisen X kallas designmatrisen, vektorn $\mathbf{y}$ kallas observationsvektorn, vektorn $\boldsymbol{\beta}$ kallas parametervektorn. Minstakvadratlösning hittas som lösning till normalekvationen

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

Matrisen $\left(X^T X\right)$ har i det fallet storlek 2×2 och normalekvationen kan lösas med hjälp av Cramers formeln.

EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2, 1)$, $(5, 2)$, $(7, 3)$, and $(8, 3)$.

SOLUTION Use the x-coordinates of the data to build the design matrix X in (1) and the y-coordinates to build the observation vector $\mathbf{y}$:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, obtain the normal equations (with the new notation):

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

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1}\begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix}\begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14}x$$

Exempel

Betrakta experimentälla data som ligger inte längs en rak linje, men längs en krökt linje som kan approximeras med en parabel $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

I det fallet framställs önskade approximativa relationen mellan experimentälla värden av x och y som ett (olösbart) system ekvationer

Experimentella argument	Linjär approximation	Experimentälla funktions värden
x_1	$\beta_0 + \beta_1 x_1 + \beta_2 x_1^2$	$= y_1$
x_2	$\beta_0 + \beta_1 x_2 + \beta_2 x_2^2$	$= y_2$
x_m	$\beta_0 + \beta_1 x_m + \beta_2 x_m^2$	$= y_m$

Vi söker igen en minstakvadratlösning till ekvationen

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

där

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

som minimerar summan av **residualer** $\epsilon_k = (\beta_0 + \beta_1 x_2 + \beta_2 x_k^2 - y_k)$ i kvadrat.

Man kan betrakta flera olika approximationer av experimentalla data med modeller som är linjära kombinationer av olika lämpliga funktioner med ett inte för stort antal parametrar $\boldsymbol{\beta}$. Matrisen $X^T X$ i normalekvationen har i det fallet storlek 3×3. Kolonnerna i X är linjärt oberoende. Detta medfär att normalekvationen

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

har en entydig lösning.

Exempel då kolonner i matrisen A är ortogonala.

EXAMPLE 4 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

SOLUTION Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 \tag{5}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

Now that $\hat{\mathbf{b}}$ is known, we can solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. But this is trivial, since we already know what weights to place on the columns of A to produce $\hat{\mathbf{b}}$. It is clear from (5) that

$$\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$

Att anpassa experimentälla data med flera faktorer.

Det kan uppstå situationer då experimentälla data är funktioner som beror på flera faktorer.

Betrakta fallet då vi vill approximera experimentella data $\mathbf{y}$ som beror på värdena av två oberoende faktorer $\mathbf{u}$ och $\mathbf{v}$. Vi väljer igen en linjär approximation:

$$y = \beta_0 + \beta_1 u + \beta_3 v$$

men det kunde vara mera komplicerade funktioner av u och v.

Första experimentella faktorn	Andraa experimentella faktorn	Linjär approximation	Experimentä
u_1	v_1	$\beta_0 + \beta_1 u_1 + \beta_2 v_1$	
u_2	v_2	$\beta_0 + \beta_1 u_2 + \beta_2 v_2$	
u_m	v_m	$\beta_0 + \beta_1 u_m + \beta_2 v_m$	

Vi söker igen en minstakvadratlösning till ekvationen

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

där

$$X = \begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_m & v_m \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

som minimerar summan av **residualer** $\epsilon_k = (\beta_0 + \beta_1 u_k + \beta_2 v_k - y_k)$ i kvadrat.

Minstakvadratlösning satisfierar normala ekvationen

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

där matrisen $X^T X$ har storlek 3×3.

EXAMPLE 7 The matrix

$$U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}$$

is an orthogonal matrix because it is square and because its columns are orthonormal, by Example 5. Verify that the rows are orthonormal, too! ■

EXAMPLE 4 The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. 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 By the Best Approximation Theorem, the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$$\hat{\mathbf{y}} = \frac{15}{30}\mathbf{u}_1 + \frac{-21}{6}\mathbf{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = 3^2 + 6^2 = 45$$

The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$. ■