Objectives

- Perform Orthogonal Decomposition of a vector onto a subspace of $\mathbb{R}^n$
- Obtain an orthogonal basis from a basis for a subspace of $\mathbb{R}^n$
- Obtain a QR factorization for a matrix
- Find the least-squares solution to the overdetermined system $A\mathbf{x} = \mathbf{b}$
- Obtain the least-squares regression line/curve that best fits data points

6.3 Orthogonal Projections

Orthogonal projection

The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is given by,

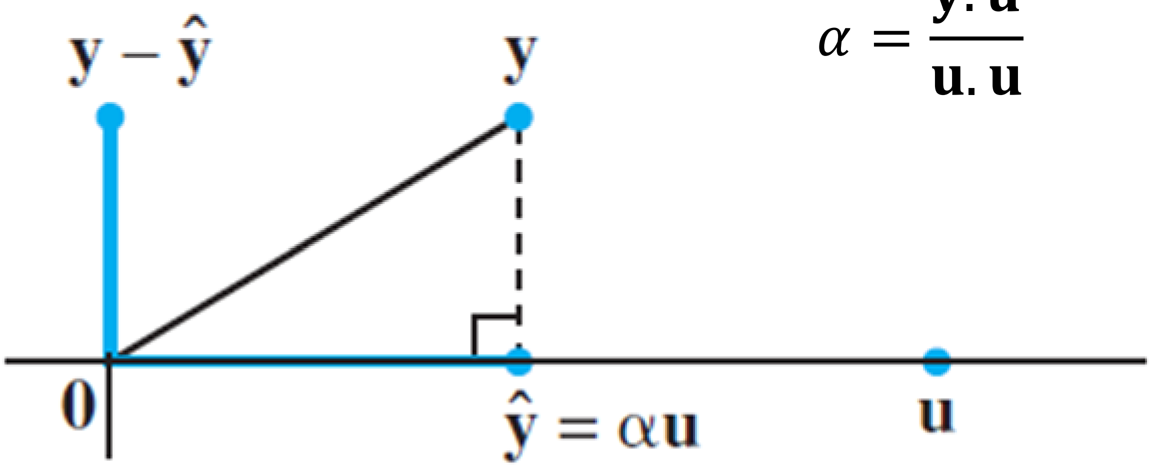
$$\hat{\mathbf{y}} = proj_L \mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$$

where $L = \text{Span}\{\mathbf{u}\}$

$$(\mathbf{y} - \alpha\mathbf{u})\cdot(\mathbf{u}) = 0$$

$$\mathbf{y}\cdot\mathbf{u} - \alpha\mathbf{u}.\mathbf{u} = 0$$

$$\alpha = \frac{\mathbf{y}.\mathbf{u}}{\mathbf{u}.\mathbf{u}}$$



6.3 Orthogonal Projections

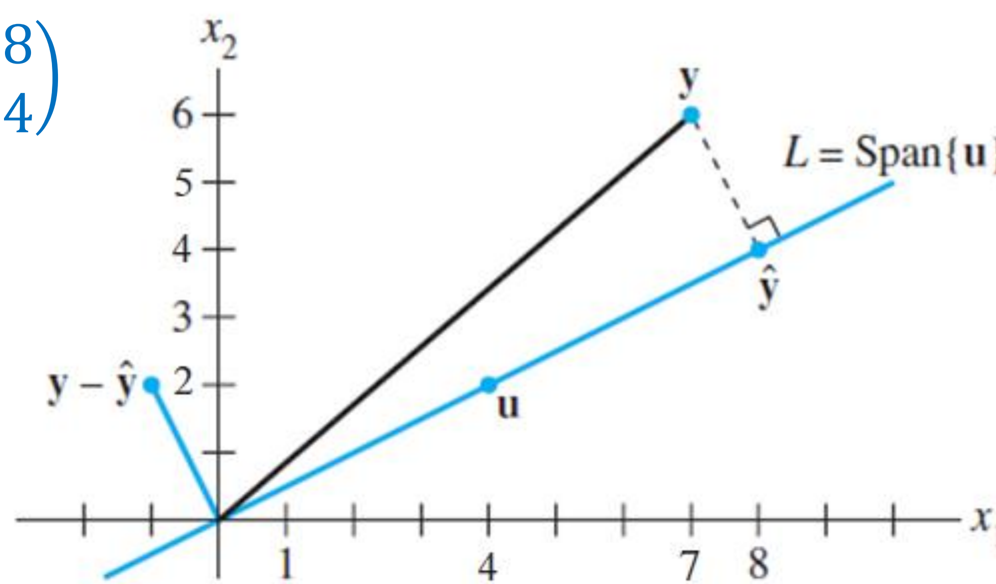
Ex. Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$, then write $\mathbf{y}$ as the sum of 2 vectors, one in $\text{Span}\{\mathbf{u}\}$ and the other orthogonal to $\mathbf{u}$

$$\mathbf{y} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}, \mathbf{u} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

$$\hat{\mathbf{y}} = proj_L \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} = 2\mathbf{u} = \begin{pmatrix} 8 \\ 4 \end{pmatrix}$$

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

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$



6.3 Orthogonal Projections

Ex. Express $\mathbf{y}$ as a linear combination of $\mathbf{u}_1, \mathbf{u}_2$

$$\mathbf{y} = \begin{pmatrix} -0.5 \\ 5.5 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -5 \\ 5 \end{pmatrix}$$

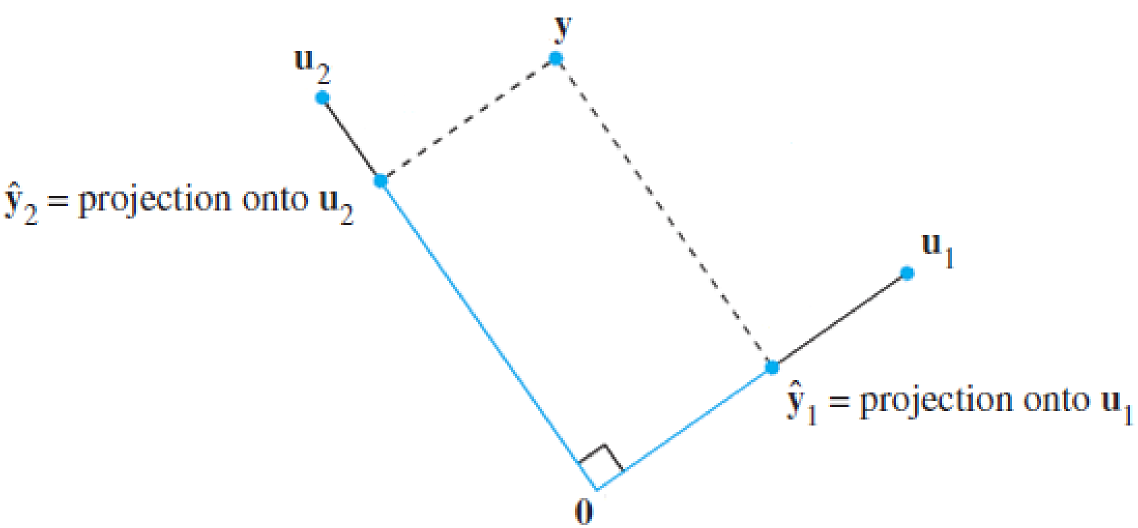
$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \qquad \mathbf{y} \cdot \mathbf{u}_1 = c_1\mathbf{u}_1 . \mathbf{u}_1$$

$$c_1 = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} = 0.625$$

$$c_2 = \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} = 0.6$$

$$\mathbf{y} = 0.625\mathbf{u}_1 + 0.6\mathbf{u}_2$$

$$\mathbf{y} = \hat{\mathbf{y}}_1 + \hat{\mathbf{y}}_2$$



6.3 Orthogonal Projections

Orthogonal Decomposition

Let W be a subspace of $\mathbb{R}^n$, then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

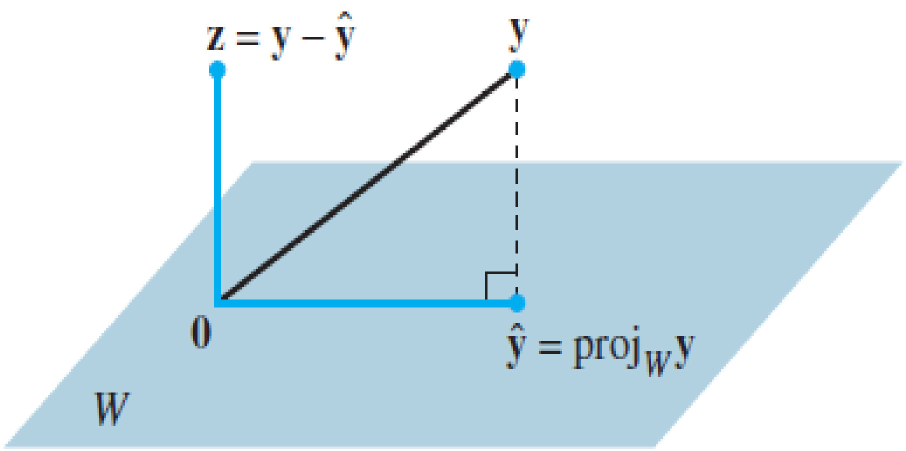
$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \text{ where } \hat{\mathbf{y}} \in W, \mathbf{z} \in W^{\perp}$$

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \cdots \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p$$

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

$S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_p\}$ is an orthogonal basis for W

Note that: $(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_i = 0$



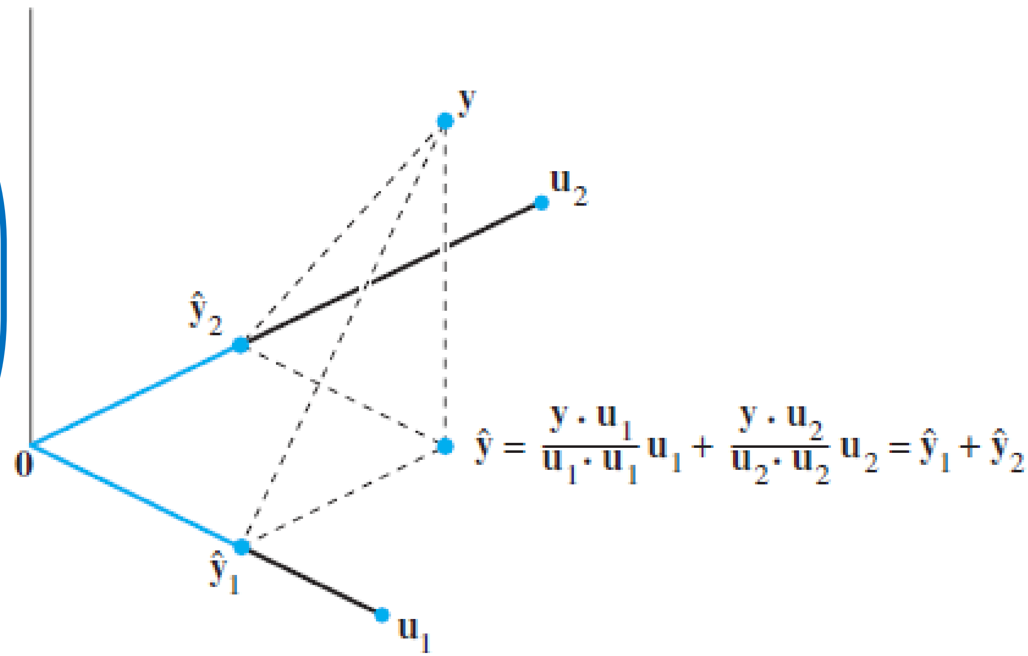
6.3 Orthogonal Projections

Ex. Write $\mathbf{y}$ as the sum of a vector in W and a vector in $W^{\perp}$

$$\mathbf{y} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$$

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

$$\hat{\mathbf{y}} = \begin{pmatrix} \frac{-2}{5} \\ 2 \\ \frac{1}{5} \end{pmatrix}, \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{pmatrix} \frac{7}{5} \\ 0 \\ \frac{14}{5} \end{pmatrix}$$



6.3 Orthogonal Projections

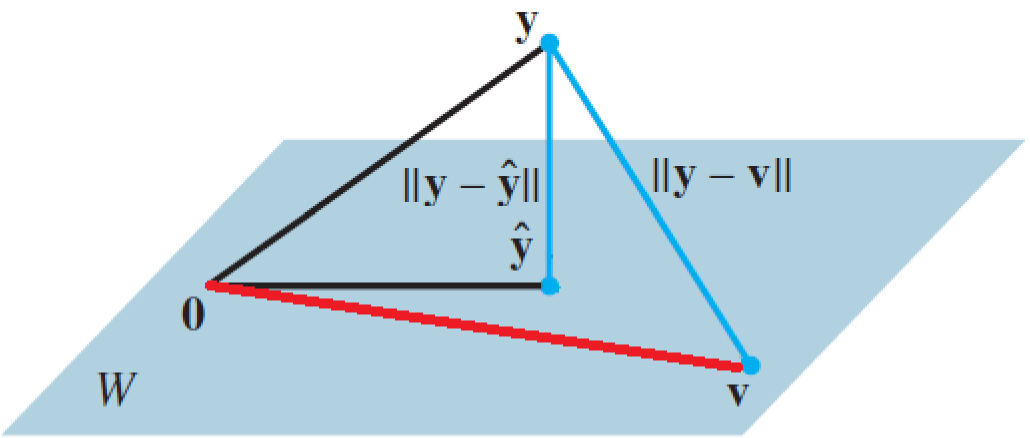
Best Approximation Theorem

Let W be a subspace of R^n, $\mathbf{y} \in R^n$, for all $\mathbf{v} \in W$,

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

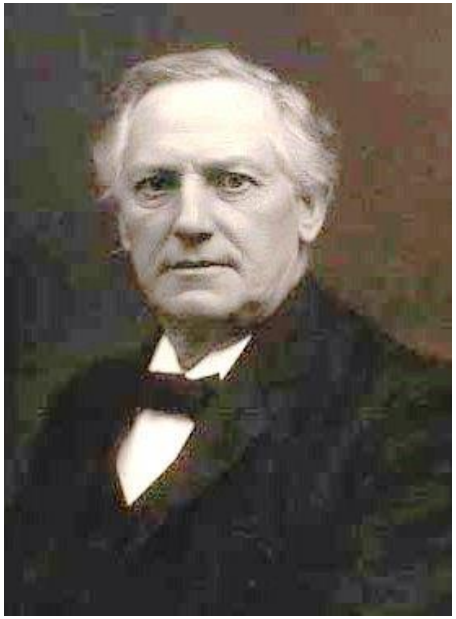
$\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto W

i.e. The closest vector to $\mathbf{y}$ in W is $\hat{\mathbf{y}}$



6.4 Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_p\}$ for a subspace W in R^n, an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_p\}$ for W can be obtained



Jørgen Pedersen Gram
Danish Mathematician
1850 - 1916

Erhard Schmidt
German Mathematician
1876 - 1959

6.4 Gram-Schmidt Process

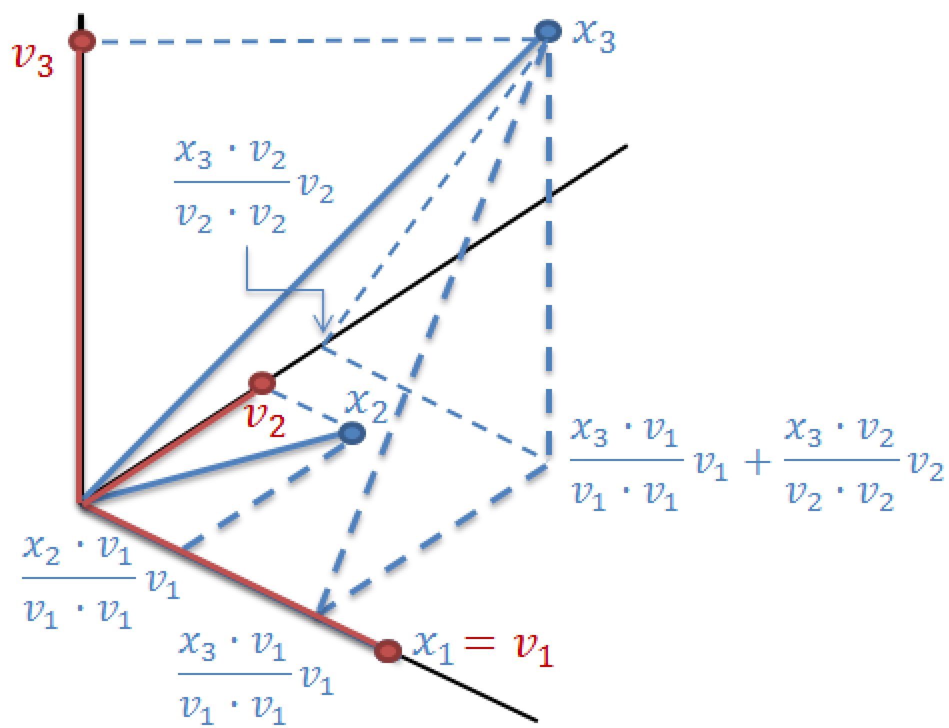
Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_p\}$ for a subspace W in R^n, an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_p\}$ for W can be obtained

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

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

$$\vdots$$



6.4 Gram-Schmidt Process

Ex. Construct an orthonormal basis for R^3 from the vectors,

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

Vectors are not orthogonal, apply Gram–Schmidt,

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

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - \frac{1+2}{1+1}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}$$

6.4 Gram-Schmidt Process

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix},$$

$$\mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\mathbf{v}_2 = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}$$

$$\mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \frac{1}{1+1}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{\frac{1}{2}}{\frac{1}{4}+\frac{1}{4}}\begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

6.4 Gram-Schmidt Process

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

To get an orthonormal set, just divide each vector by its length,

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \qquad \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \qquad \mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

6.4 Gram-Schmidt Process

Ex. Find the orthogonal projection of $\mathbf{y}$ onto W

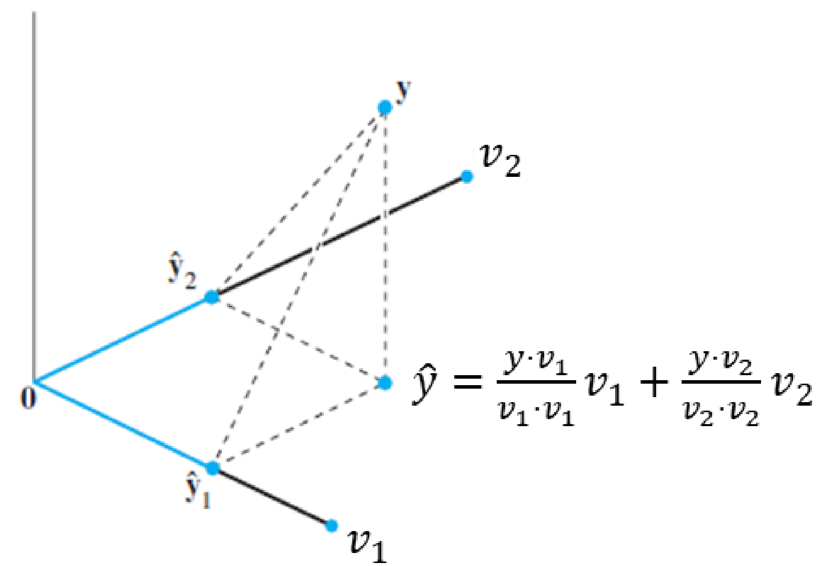
$$\mathbf{y}=\begin{pmatrix}1\\2\\3\end{pmatrix},\mathbf{x}_1=\begin{pmatrix}2\\5\\-1\end{pmatrix},\mathbf{x}_2=\begin{pmatrix}0\\6\\0\end{pmatrix},W=\operatorname{Span}\{\mathbf{x}_1,\mathbf{x}_2\}$$

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

$$\mathbf{v}_2=\mathbf{x}_2-\frac{\mathbf{x}_2\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1$$

$$=\begin{pmatrix}0\\6\\0\end{pmatrix}-\frac{30}{30}\begin{pmatrix}2\\5\\-1\end{pmatrix}=\begin{pmatrix}-2\\1\\1\end{pmatrix}$$

$$\hat{\mathbf{y}}=\frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1+\frac{\mathbf{y}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2=\begin{pmatrix}-\frac{2}{5}\\2\\\frac{1}{5}\end{pmatrix}$$



6.4 Gram-Schmidt Process

QR Factorization

If A is an $m \times n$ matrix with linearly independent columns

$$A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]$$

Applying Gram–Schmidt orthogonalization to columns of A,

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

$$\mathbf{v}_2 = \mathbf{a}_2 - \frac{\mathbf{a}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

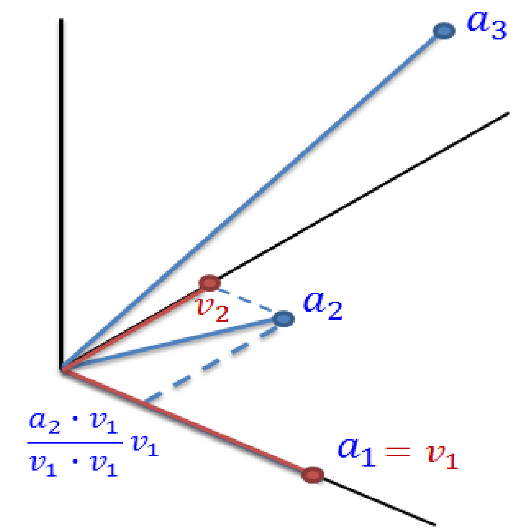
$$\mathbf{v}_3 = \mathbf{a}_3 - \frac{\mathbf{a}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{a}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

$$\vdots$$

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|}, \ldots$$

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

$$\mathbf{a}_2 = r_{12}\mathbf{u}_1 + \|\mathbf{v}_2\|\mathbf{u}_2$$



6.4 Gram-Schmidt Process

QR Factorization

$$\mathbf{a}_1 = \|\mathbf{v}_1\|\mathbf{u}_1$$
$$\mathbf{a}_2 = r_{12}\mathbf{u}_1 + \|\mathbf{v}_2\|\mathbf{u}_2$$

$$\underbrace{[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]}_{A} = \underbrace{[\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n]}_{Q} \underbrace{\begin{bmatrix} \|\mathbf{v}_1\| & r_{12} & \cdots & r_{1n} \\ 0 & \|\mathbf{v}_2\| & \cdots & r_{2n} \\ 0 & 0 & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\mathbf{v}_n\| \end{bmatrix}}_{R}$$

Note that:

$A = QR$

$Q^T Q = I_n$

$$\therefore R = Q^T A$$

R: is an invertible upper triangular matrix

6.4 Gram-Schmidt Process

Ex. Find a QR factorization for,

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 3 & 7 \\ 1 & -1 & -4 \\ 1 & -1 & 2 \end{pmatrix}$$

Applying Gram–Schmidt to column of A leads to:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ 2 \\ -2 \\ -2 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -3 \\ 3 \\ -3 \\ 3 \end{pmatrix}$$

6.4 Gram-Schmidt Process

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 3 & 7 \\ 1 & -1 & -4 \\ 1 & -1 & 2 \end{pmatrix}$$

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ 2 \\ -2 \\ -2 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -3 \\ 3 \\ -3 \\ 3 \end{pmatrix}$$

$$Q = \frac{1}{2}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \qquad R = Q^T A = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$$

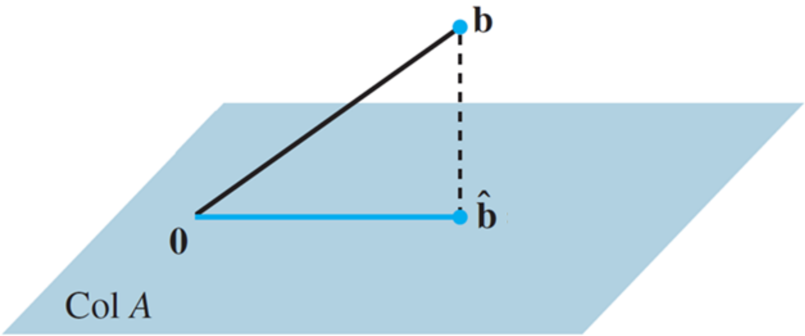
6.5 Least-Squares Problems

- $A\mathbf{x} = \mathbf{b}$ has a solution if

$$x_1\mathbf{a_1} + x_2\mathbf{a_2} \dots x_n\mathbf{a}_n = \mathbf{b}$$
$$\mathbf{b} \in \text{Col } A$$

- $A\mathbf{x} = \mathbf{b}$ has no solution if

$$\mathbf{b} \notin \text{Col } A$$



6.5 Least-Squares Problems

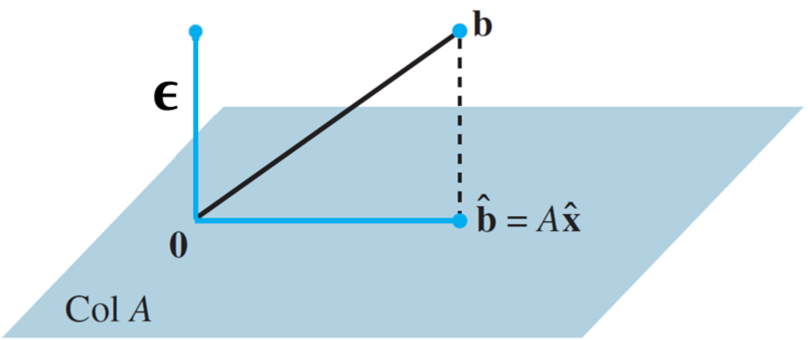
For $A_{m\times n}$ and $\mathbf{b} \in R^m$, a least–squares solution for $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in R^n such that

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

where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto $\mathrm{Col}\ A$

Residual vector

$$\mathbf{b} = \hat{\mathbf{b}} + \boldsymbol{\epsilon} = A\hat{\mathbf{x}} + \boldsymbol{\epsilon}$$



6.5 Least-Squares Problems

$$\mathbf{a}_i.\left(\mathbf{b}-\hat{\mathbf{b}}\right)=0$$

$$\mathbf{a}_1^T\left(\mathbf{b}-\hat{\mathbf{b}}\right)=0$$
$$\mathbf{a}_2^T\left(\mathbf{b}-\hat{\mathbf{b}}\right)=0$$
$$A^T\left(\mathbf{b}-\hat{\mathbf{b}}\right)=\mathbf{0}$$
$$A^T(\mathbf{b}-A\hat{\mathbf{x}})=\mathbf{0}$$

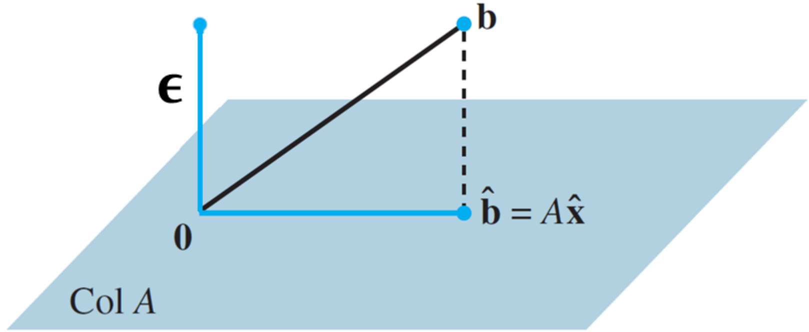
$A^TA\hat{\mathbf{x}}=A^T\mathbf{b}$ Normal Eqs.

If A has linearly independent columns,

$$\hat{\mathbf{x}}=\left(A^TA\right)^{-1}A^T\mathbf{b}=A^+\mathbf{b}$$

where $A^+=\left(A^TA\right)^{-1}A^T$ is called pseudo–inverse of A (Moore–Penrose inverse)

$$A^+A=I_n$$



6.5 Least-Squares Problems

Ex. Find the least-squares solution of the overdetermined system $A\mathbf{x} = \mathbf{b}$ and find the projection of $\mathbf{b}$ onto $\mathrm{Col}\,A$ where,

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 3 & 7 \\ 1 & -1 & -4 \\ 1 & -1 & 2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 12 \end{pmatrix}$$

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

$$\begin{pmatrix} 4 & 4 & 6 \\ 4 & 20 & 26 \\ 6 & 26 & 70 \end{pmatrix}\hat{\mathbf{x}} = \begin{pmatrix} 16 \\ -4 \\ 35 \end{pmatrix} \Rightarrow \hat{\mathbf{x}} = \begin{pmatrix} 5 \\ -2.5 \\ 1 \end{pmatrix} \qquad \hat{\mathbf{b}} = A\hat{\mathbf{x}} = \begin{pmatrix} -1.5 \\ 4.5 \\ 3.5 \\ 9.5 \end{pmatrix}$$

6.5 Least-Squares Problems

Using QR factorization for A leads to:

$$Q = \frac{1}{2}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \quad R = Q^T A = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 3 & 7 \\ 1 & -1 & -4 \\ 1 & -1 & 2 \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 12 \end{pmatrix}$$

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b} \;\Rightarrow R\hat{\mathbf{x}} = Q^T\mathbf{b}$$

$$\begin{pmatrix} 2 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}\hat{\mathbf{x}} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 12 \end{pmatrix}$$

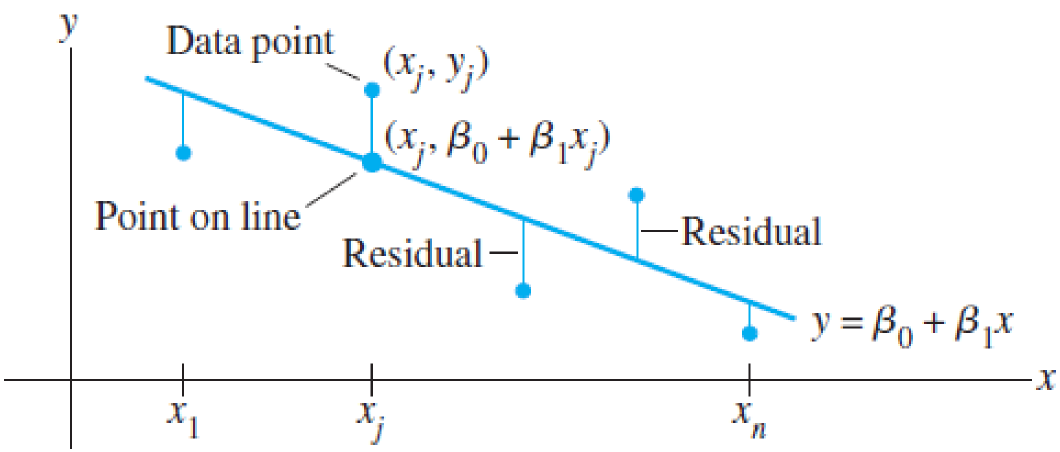
$$\hat{\mathbf{x}} = \begin{pmatrix} 5 \\ -2.5 \\ 1 \end{pmatrix}$$

6.6 Machine Learning and Linear Models

Least-Squares Lines

Given a set of data points $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$
It is required to find the line $y = \beta_0 + \beta_1 x$ that best fits the data points

Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	$=$	y_1
$\beta_0 + \beta_1 x_2$	$=$	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	$=$	y_n



6.6 Machine Learning and Linear Models

Least-Squares Lines

Given a set of data points $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$
It is required to find the line $y = \beta_0 + \beta_1 x$ that best fits the data points

Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	$=$	y_1
$\beta_0 + \beta_1 x_2$	$=$	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	$=$	y_n

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

X Design Matrix

$\boldsymbol{\beta}$

$\mathbf{y}$ Observation Vector

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

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

6.6 Machine Learning and Linear Models

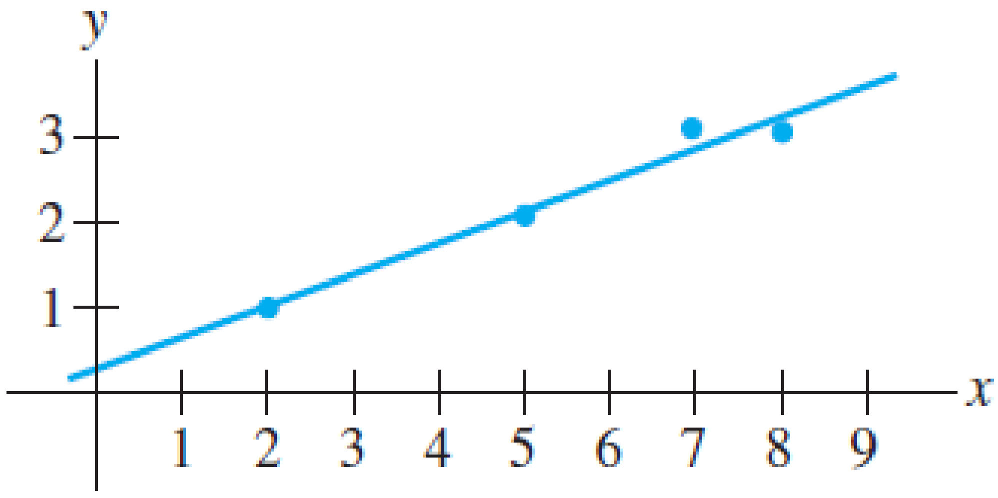
Ex. Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points (2,1),(5,2),(7,3),(8,3)

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

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

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

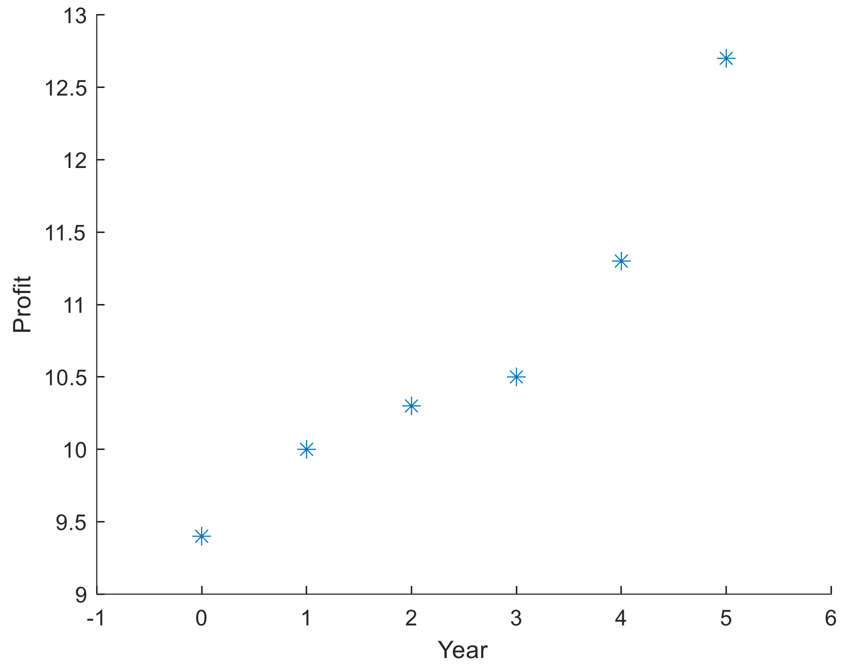
$$y = \frac{2}{7} + \frac{5}{14} x$$



6.6 Machine Learning and Linear Models

Ex. The net profits (in billions of dollars) for Microsoft from 2000 to 2005 are shown in the table,

x (Year 20−)	y (Profit)
00	9.4
01	10.0
02	10.3
03	10.5
04	11.3
05	12.7



6.6 Machine Learning and Linear Models

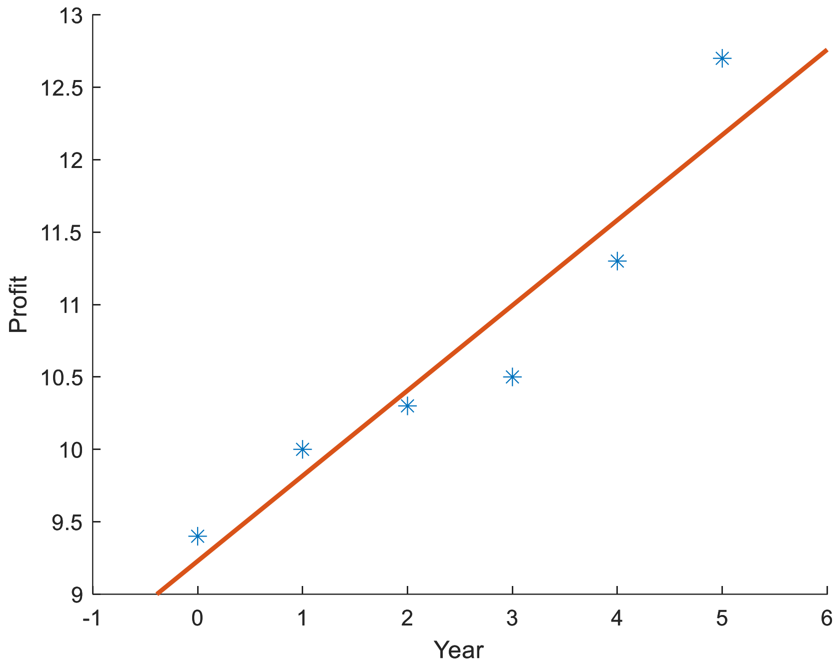
Find the least squares regression line that best fits the data

$$y = \beta_0 + \beta_1 x$$
$$\beta_0 = 9.23$$
$$\beta_1 = 0.59$$

Predict the profit in 2006

$$y(6) = 9.23 + 0.59(6)$$
$$= \$12.77 \text{ Billion}$$

Actual was $12.60 Billion



6.6 Machine Learning and Linear Models

Least-Squares Polynomials

Given a set of data points $(x_j, y_j), j = 1 \cdots n$

Required to find

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m$$

that best fits the data points

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

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

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

6.6 Machine Learning and Linear Models

Ex. Use least-squares regression to find the equation of the parabola $y = \beta_0 + \beta_1 x + \beta_2 x^2$ that best fits the data points (1,1),(2,5),(4,14),(5,22)

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

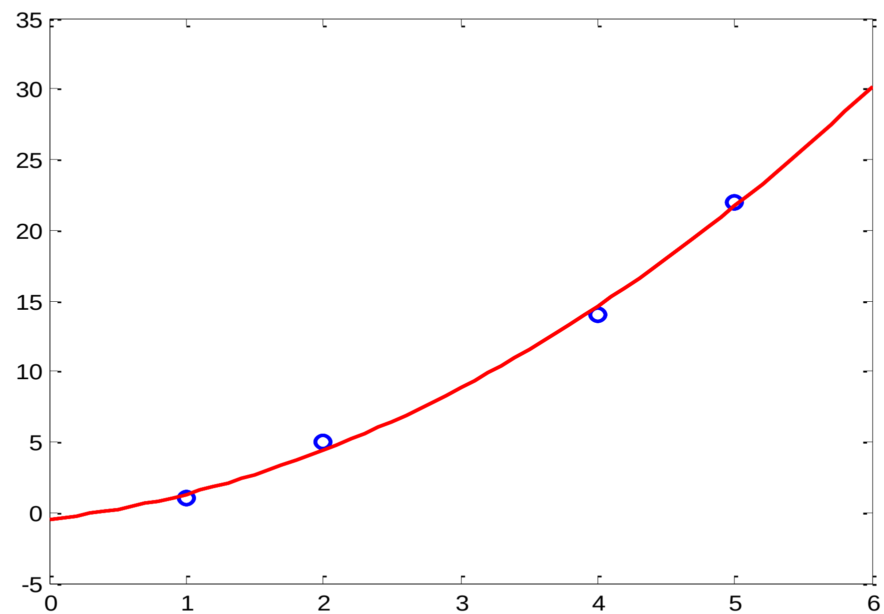
$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 1 \\ 5 \\ 14 \\ 22 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 12 & 46 \\ 12 & 46 & 198 \\ 46 & 198 & 898 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 42 \\ 177 \\ 795 \end{pmatrix}$$

$$y = \frac{-14}{30} + \frac{33}{30}x + \frac{20}{30}x^2$$

6.6 Machine Learning and Linear Models

Ex. Use least-squares regression to find the equation of the parabola $y = \beta_0 + \beta_1 x + \beta_2 x^2$ that best fits the data points (1,1),(2,5),(4,14),(5,22)



6.6 Machine Learning and Linear Models

Given a set of data points $(x_j, y_j), j = 1 \cdots n$

Required to find

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \beta_2 f_2(x) + \cdots + \beta_m f_m(x)$$

that best fits the data points

$$X = \begin{pmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_m(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_m(x_n) \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

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

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

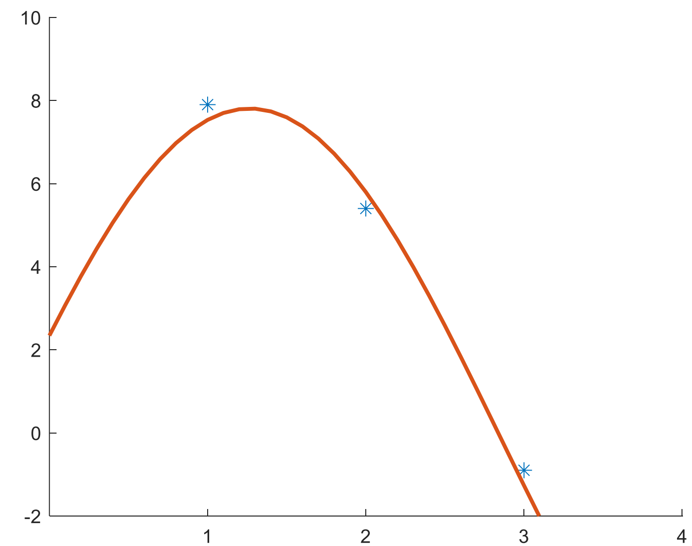
6.6 Machine Learning and Linear Models

Ex. A certain experiment produces the data (1,7.9), (2,5.4), (3,–0.9). Find the model that produces a least–squares fit of these points by a function of the form $y = \beta_0 \cos(x) + \beta_1 \sin(x)$

$$X = \begin{pmatrix} \cos(1) & \sin(1) \\ \cos(2) & \sin(2) \\ \cos(3) & \sin(3) \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 7.9 \\ 5.4 \\ -0.9 \end{pmatrix}$$

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

$$\begin{pmatrix} 1.44 & -0.06 \\ -0.06 & 1.55 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 2.91 \\ 11.43 \end{pmatrix}$$



$$y = 2.34 \cos(x) + 7.45 \sin(x)$$