CS 260: Foundations of Data Science

Prof. Sara Mathieson

Fall 2021

HAVERFORD COLLEGE

Admin

- **Lab 2** due TONIGHT!
  - Lab 3 posted later today

- **Office hours TODAY**: 3:30-5pm in H204

- **TA hours TONIGHT**: 6:30-8:30pm in H110

- **Note-taker**: Tazkia

Outline for September 14

- Recap simple (i.e. $p=1$) linear regression

- Introduction to applied linear algebra

- Multiple linear regression

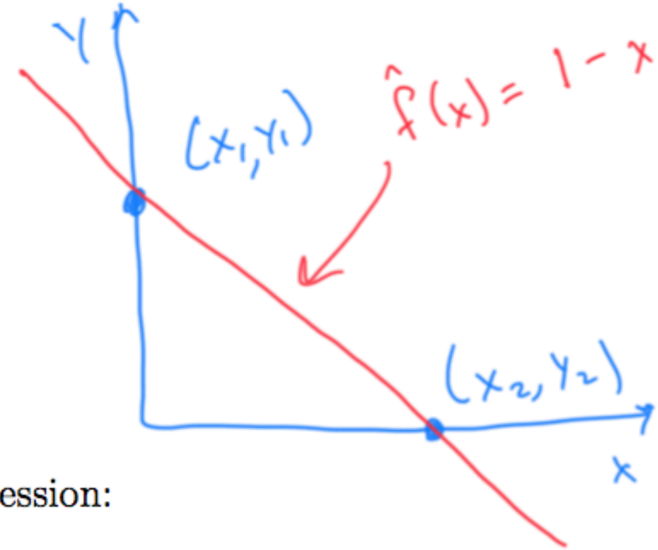
- Analytic solution to multiple linear regression

Outline for September 14

- Recap simple (i.e. p=1) linear regression
- Introduction to applied linear algebra
- Multiple linear regression
- Analytic solution to multiple linear regression

Handout 4

1. *Toy example.* Let $n = 2$ and $p = 1$, with the following data (we will omit the first column of 1's in simple linear regression):



$$\boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

(a) Plot these two points – what should $\hat{w}_0$ and $\hat{w}_1$ be?

$\hat{w}_0 = 1$

$\hat{w}_1 = -1$

(b) This week we derived the solution for simple linear regression:

note: $\bar{x} = \frac{1}{2}$

$\bar{y} = \frac{1}{2}$

$$\hat{w}_1 = \frac{\text{Cov}(\boldsymbol{x}, \boldsymbol{y})}{\text{Var}(\boldsymbol{x})} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Use these equations to compute $\hat{w}_0$ and $\hat{w}_1$ and verify your answer to (a).

$$\hat{w}_1 = \frac{\cancel{\tfrac{1}{2}}\left[(1-\tfrac{1}{2})(0-\tfrac{1}{2}) + (0-\tfrac{1}{2})(1-\tfrac{1}{2})\right]}{\cancel{\tfrac{1}{2}}\left[(1-\tfrac{1}{2})^2 + (0-\tfrac{1}{2})^2\right]}$$

$$= \frac{-\frac{1}{4} - \frac{1}{4}}{\frac{1}{4} + \frac{1}{4}} \Rightarrow \boxed{\hat{w}_1 = -1}$$

$$\Rightarrow \hat{w}_0 = \frac{1}{2} - (-1)\frac{1}{2}$$

$$\Rightarrow \boxed{\hat{w}_0 = 1}$$

Outline for September 14

- Recap simple (i.e. p=1) linear regression
- Introduction to applied linear algebra
- Multiple linear regression
- Analytic solution to multiple linear regression

Vectors

- Vector magnitude

$$\mathbf{v} = \left[\begin{array}{c} x \\ y \end{array} \right], \quad \text{then} \quad |\mathbf{v}| = \sqrt{x^2 + y^2}.$$

- Different ways to write a vector

$$\mathbf{v} = \left[\begin{array}{c} x \\ y \end{array} \right] = \left[\begin{array}{cc} x & y \end{array} \right]^{\mathrm{T}}$$

- Vector dot product

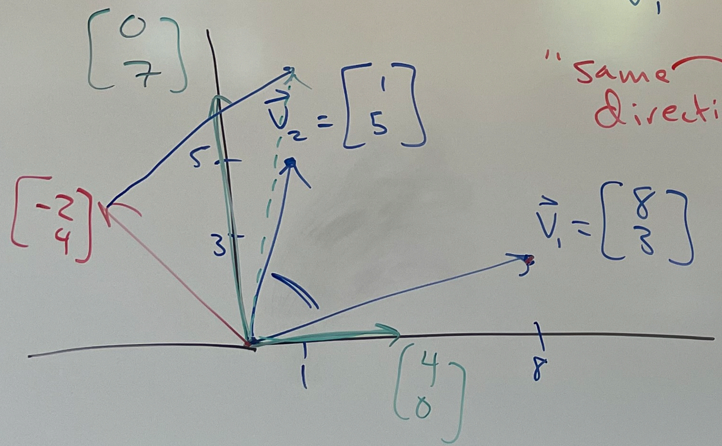
PLEASE LEAVE COMPUTERS ON

do not erase

$p=1$: "simple" linear regression

$$h_{\vec{w}}(x) = w_0 \boxed{1} + w_1 x = \vec{w} \cdot \vec{x} \leftarrow \begin{bmatrix}1\\x\end{bmatrix}$$

Dot product



$$\vec{v}_1 \cdot \vec{v}_2 = 8 \cdot 1 + 3 \cdot 5 = \boxed{23}$$

"same direction"

always a scalar

$$\begin{bmatrix}4\\0\end{bmatrix} \cdot \begin{bmatrix}0\\7\end{bmatrix} = 4 \cdot 0 + 0 \cdot 7 = \boxed{0}$$

"perpendicular"

$$\begin{bmatrix}8\\3\end{bmatrix} \cdot \begin{bmatrix}-2\\4\end{bmatrix} = -16 + 12 = \boxed{-4}$$

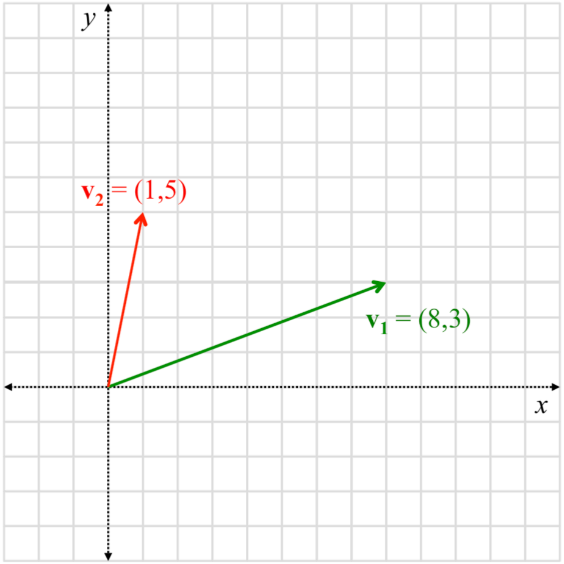
"opposite direction"

$$\vec{a} = \begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}, \quad \vec{b} = \begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}$$

$$\Rightarrow \boxed{\vec{a} \cdot \vec{b} = \sum_{i=1}^{n} a_i b_i}$$

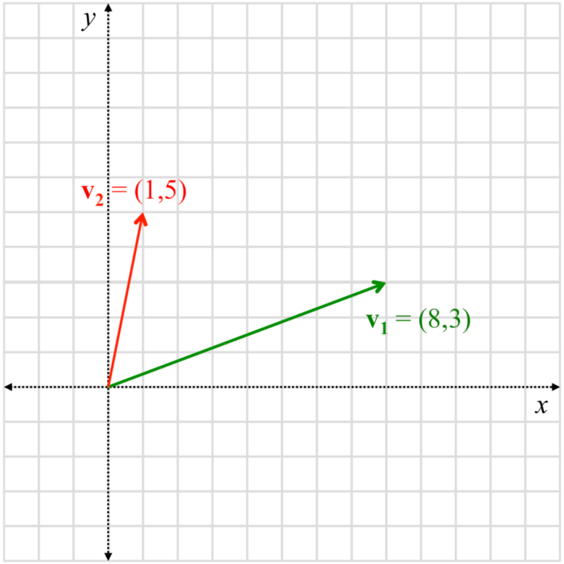
Vector Addition

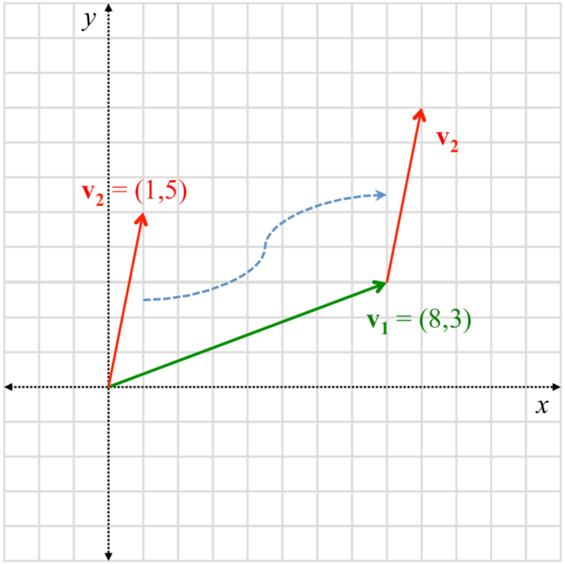
$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 8+1 \\ 3+5 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$



Vector Addition

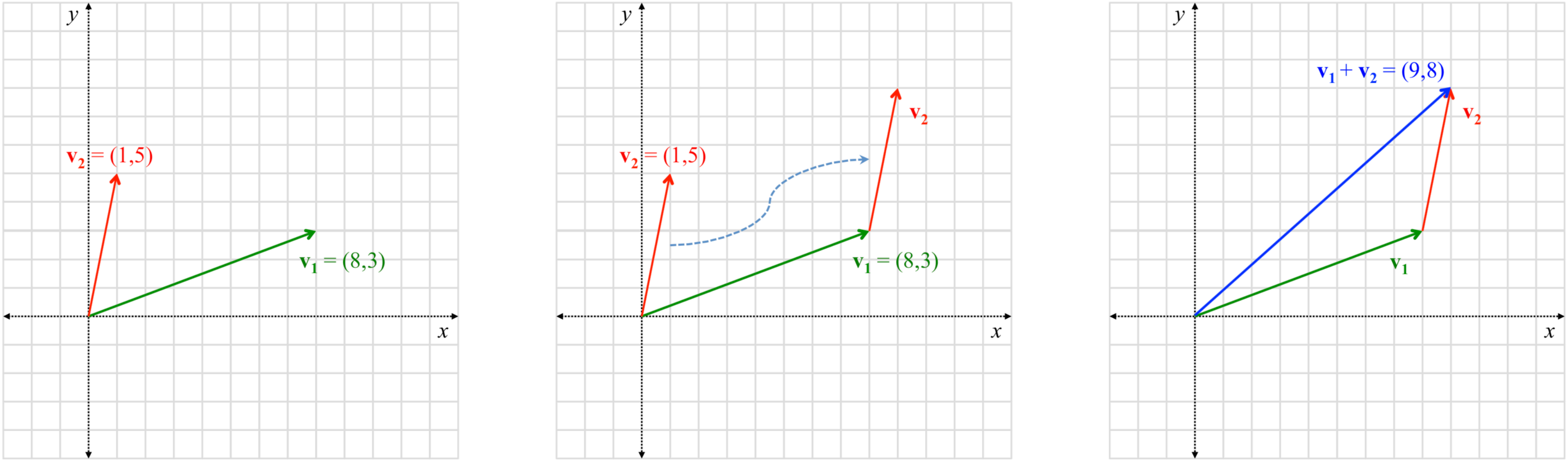
$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 8+1 \\ 3+5 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$





Vector Addition

$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 8+1 \\ 3+5 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$



Matrices

- Matrix addition (must be exactly the same dimension!)

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}+\begin{bmatrix} e & f \\ g & h \end{bmatrix}=\begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

Matrix Multiplication

- inner dimensions must match
- If A.shape = (m, n) and B.shape = (n, p), then AB.shape = (m,p)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & \\ & \end{bmatrix}$$

dot product of row 1 and col 1

Matrix Multiplication

- inner dimensions must match
- If A.shape = (m, n) and B.shape = (n, p), then AB.shape = (m,p)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & \\ & \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & \boxed{af + bh} \\ & \end{bmatrix}$$

dot product of row 1 and col 2

Matrix Multiplication

- inner dimensions must match
- If A.shape = (m, n) and B.shape = (n, p), then AB.shape = (m,p)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & \\ & \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ & \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

Outline for September 14

- Recap simple (i.e. p=1) linear regression
- Introduction to applied linear algebra
- Multiple linear regression
- Analytic solution to multiple linear regression

Lab 3: USA Housing data

Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
79545.45857	5.682861322	7.009188143	4.09	23086.8005	1059033.558
79248.64245	6.002899808	6.730821019	3.09	40173.07217	1505890.915
61287.06718	5.86588984	8.51272743	5.13	36882.1594	1058987.988
63345.24005	7.188236095	5.586728665	3.26	34310.24283	1260616.807
59982.19723	5.040554523	7.839387785	4.23	26354.10947	630943.4893
80175.75416	4.988407758	6.104512439	4.04	26748.42842	1068138.074
64698.46343	6.025335907	8.147759585	3.41	60828.24909	1502055.817
78394.33928	6.989779748	6.620477995	2.42	36516.35897	1573936.564
59927.66081	5.36212557	6.393120981	2.3	29387.396	798869.5328
81885.92718	4.42367179	8.167688003	6.1	40149.96575	1545154.813
80527.47208	8.093512681	5.0427468	4.1	47224.35984	1707045.722
50593.6955	4.496512793	7.467627404	4.49	34343.99189	663732.3969
39033.80924	7.671755373	7.250029317	3.1	39220.36147	1042814.098
73163.66344	6.919534825	5.993187901	2.27	32326.12314	1291331.518
69391.38018	5.344776177	8.406417715	4.37	35521.29403	1402818.21
73091.86675	5.443156467	8.517512711	4.01	23929.52405	1306674.66
79706.96306	5.067889591	8.219771123	3.12	39717.81358	1556786.6

X

y

PLEASE LEAVE COMPUTERS ON

do not erase

Multiple Linear Regression

$$X = \begin{bmatrix} 1 & \\ \vdots & \\ 1 & \leftarrow \vec{x}_i^{\,T} \rightarrow \\ \vdots & \\ 1 & \end{bmatrix}_{n \times (p+1)}$$

"fake ones"

p = # features

n = # examples

$\vec{x}_i = [\,1 \;\; x_1 \;\; x_2 \; \cdots \; x_p\,]^T$

model

$$\hat{y} = h_{\vec{w}}(\vec{x}) = w_0 + w_1 x_1 + w_2 x_2 + \cdots w_p x_p$$

$$= \vec{w} \cdot \vec{x}$$

dot product as a model!

Goal: find $\vec{w}$

$\hat{\vec{y}}$ prediction for all data

true $\vec{y}$

$$\begin{array}{rc} & \text{feature} \\ & \\ & \\ \text{example} & \\ \end{array}
\underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ & & \vdots & & \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \\ & & & & \\ & & & & x_{np} \end{bmatrix}}_{X \;\; n \times (p+1)}
\overbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}}^{\vec{w}}_{(p+1)\times 1}
=
\overbrace{\begin{bmatrix} \hat{y}_1 \\ \\ \hat{y}_i \\ \\ \end{bmatrix}}^{\hat{\vec{y}}}_{n \times 1}
\qquad
\overbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}}^{\vec{y}}$$

how close to the truth?

Outline for September 14

- Recap simple (i.e. p=1) linear regression
- Introduction to applied linear algebra
- Multiple linear regression
- Analytic solution to multiple linear regression

Cost

minimize $J(\vec{w}) = \frac{1}{2}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$

$$J(\vec{w}) = \frac{1}{2}\sum_{i=1}^{n}(\vec{w}\cdot\vec{x}_i - y_i)^2$$

$$= \frac{1}{2}(\vec{\hat{y}} - \vec{y})\cdot(\vec{\hat{y}} - \vec{y})$$

minimize

$$\boxed{J(\vec{w}) = \frac{1}{2}\overset{(a\ -\ b)}{(X\vec{w} - \vec{y})}\cdot\overset{(a\ -\ b)}{(X\vec{w} - \vec{y})}}$$

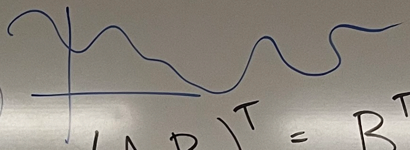
$$= \frac{1}{2}\left(\underbrace{(X\vec{w})\cdot(X\vec{w})}_{a^2} \underbrace{- 2\vec{y}\cdot(X\vec{w})}_{-2ab} + \underbrace{\vec{y}\cdot\vec{y}}_{b^2}\right)$$

— covariance of X & y

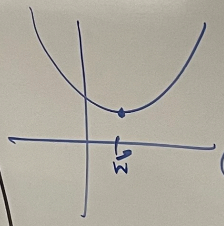
— Variance of X

$$\vec{a}\cdot\vec{b} = \sum_{i=1}^{n} a_i b_i$$

$$\vec{a}\cdot\vec{a} = \sum_{i=1}^{n} a_i^2$$



$$(AB)^T = B^T A^T$$



$$(A^T)^T = A$$

$$\vec{a}\cdot\vec{b} = \vec{a}^T\vec{b}$$

$$= \begin{bmatrix} - & \vec{a} & - \end{bmatrix}\begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix} = [\ ]$$

PLEASE LEAVE COMPUTERS ON

do not erase

take derivative and set to $\vec{0}$

$$\vec{0} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{p+1}$$

$$\boxed{w_1 = \frac{\mathrm{cov}(x,y)}{\mathrm{Var}(x)}}$$

derivative "gradient"

$$\frac{\partial J}{\partial \vec{w}} = \underbrace{\underset{(p+1)\times n}{X^T}\ \underset{n\times(p+1)}{X}\ \underset{(p+1)\times 1}{\vec{w}}}_{(p+1)\times 1} - \underbrace{\underset{(p+1)\times n}{X^T}\ \underset{n\times 1}{y}}_{(p+1)\times 1} = \vec{0}$$

best $\vec{w}$

$$A^{-1}A = I$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

<u>Solve for $\vec{w}$</u>

$$X^T X \vec{w} = X^T \vec{y}$$

$$\cancel{(X^T X)^{-1}}(X^T X \vec{w}) = (X^T X)^{-1} X^T \vec{y}$$

$$\star\ \boxed{\hat{\vec{w}} = (X^T X)^{-1} X^T \vec{y}}$$

Handout 5, #1

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 0 & 2 \\ -4 & 1 \end{bmatrix}.$$

$$\mathbf{AB} = \begin{bmatrix} -4 & 1 \\ -8 & 8 \end{bmatrix} \quad \text{and} \quad \mathbf{BA} = \begin{bmatrix} 6 & 4 \\ -1 & 6 \end{bmatrix}.$$

Handout $\boxed{5}$

#2, #3

inverse

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\vec{\omega} = \left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$