CS 260: Foundations of Data Science

Prof. Thao Nguyen

Fall 2024

HAVERFORD
COLLEGE

Materials by Sara Mathieson

Admin

- **Sit somewhere new**

- **Lab 1** grades posted on Moodle

- **Lab 2** is due tonight unless you're taking a late day

- **Lab 3** posted, due next Monday at midnight

Outline for today

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

- Introduction to applied linear algebra

- *Multiple* linear regression

- Analytic solution to multiple linear regression

Outline for today

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

- Introduction to applied linear algebra

- *Multiple* linear regression

- Analytic solution to multiple linear regression

Simple linear regression

model

$$h_{\vec{w}} = w_0 + w_1 x = \hat{y}$$

cost function

$$J(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i \right)^2$$

(model)

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

Outline for today

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

- Introduction to applied linear algebra

- *Multiple* linear regression

- Analytic solution to multiple linear regression
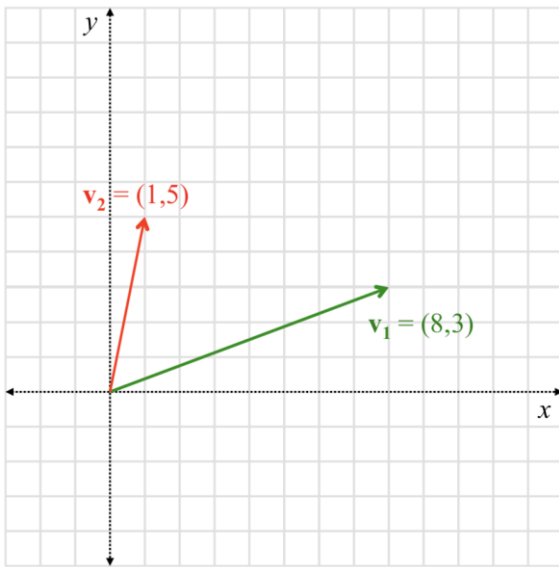
Vectors

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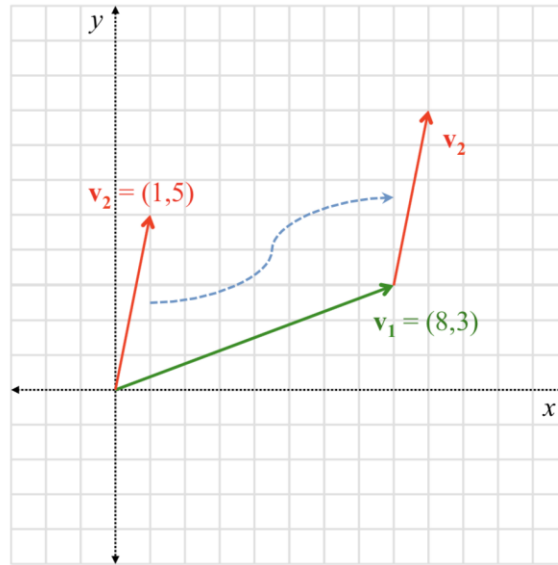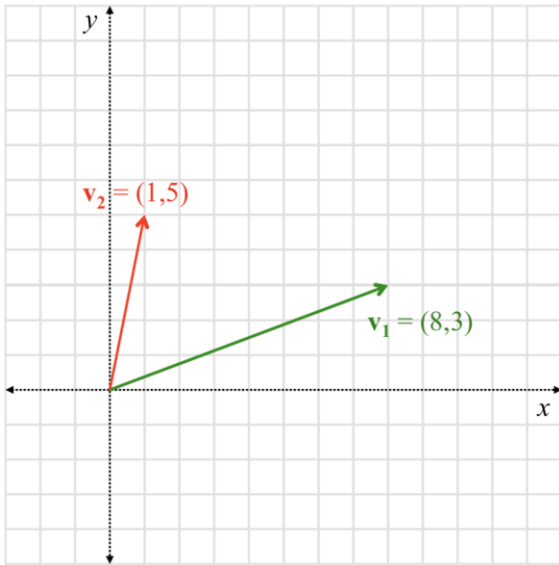
$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 8+1 \\ 3+5 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$

Vector Addition

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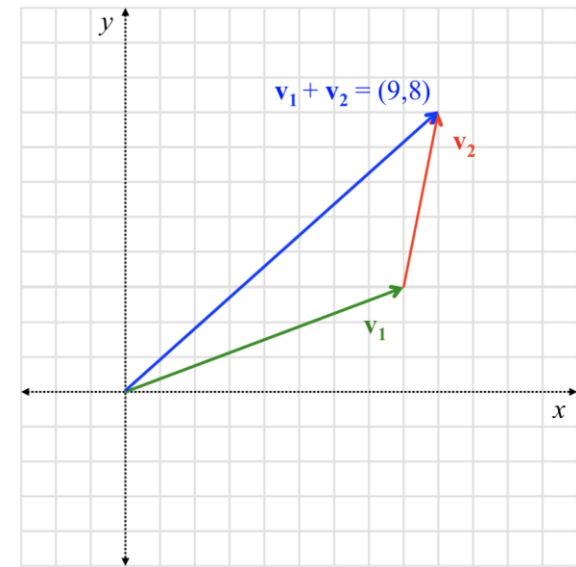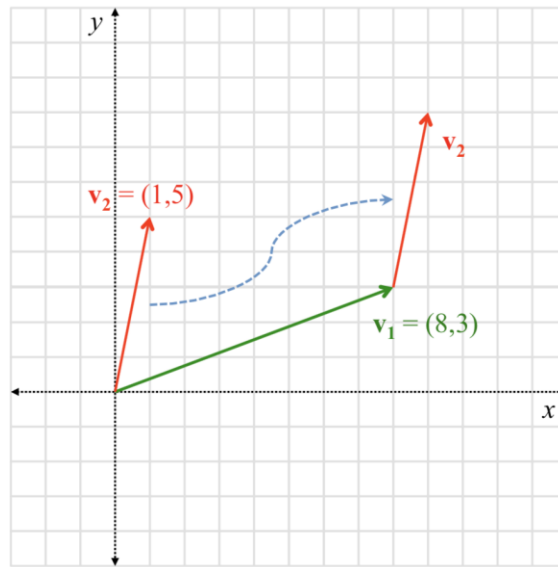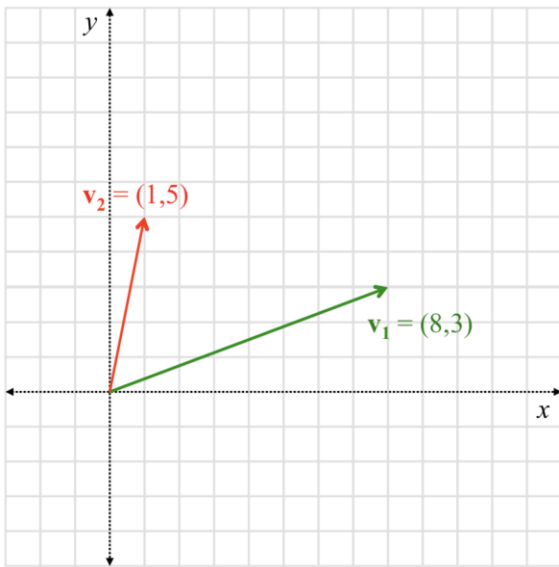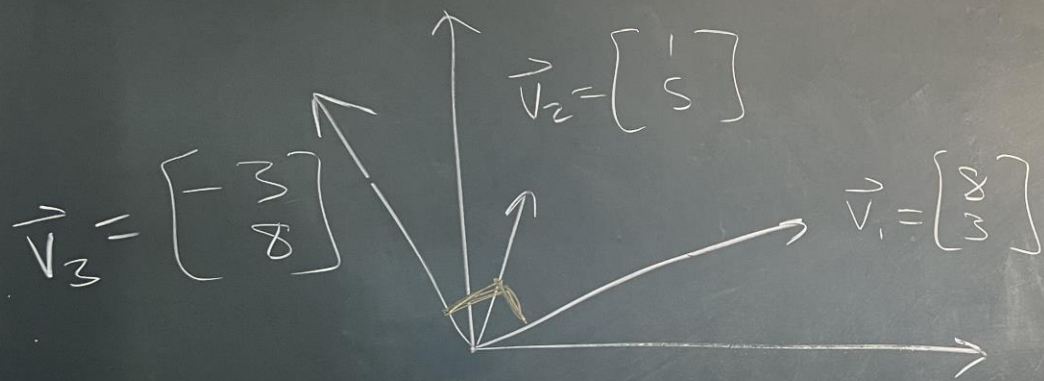
Vector Addition

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

Vector dot product

$$\vec{V_2} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

$$\vec{V_3} = \begin{bmatrix} -3 \\ 8 \end{bmatrix}$$

$$\vec{V_1} = \begin{bmatrix} 8 \\ 3 \end{bmatrix}$$

$$\vec{V_1} \circ \vec{V_3} = -3 \cdot 8 + 8 \cdot 3 = 0$$

perpendicular

$$\vec{V_1} \circ \vec{V_2} = \begin{bmatrix} 8 \\ 3 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 5 \end{bmatrix} = 8 \cdot 1 + 3 \cdot 5$$

$$= 23$$

positive → "same" direction

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)

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

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

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)

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

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

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

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

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

Matrix Transpose

$$\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \qquad \boldsymbol{A}^T = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$$

Useful note: $(\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$

Matrix Inverse

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

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

Outline for today

- 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

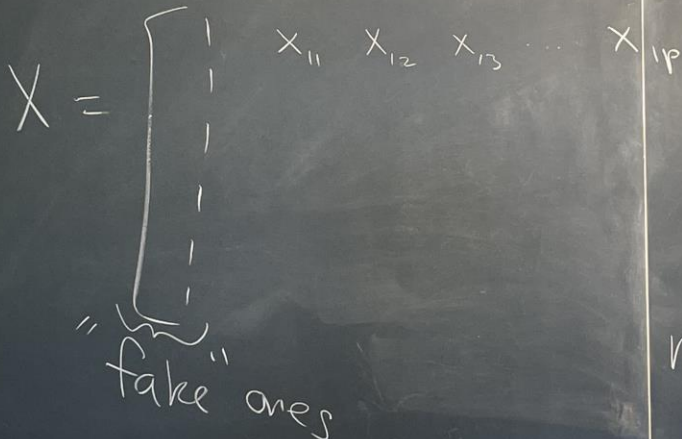X   y

Multiple Linear Regression

multiple features

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

$x_0 = 1$

want to find → dot product

$$X = \begin{bmatrix} & x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ & & & & & \\ & & & & & \\ & & & & & \end{bmatrix}$$

$\underbrace{\quad}$ "fake" ones

$n \times (p+1)$

Goal is the same

truth

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

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

★ | Minimize cost function |

$$X \begin{bmatrix} \vec{\beta} \end{bmatrix} =$$

$$\cdot (X\vec{\beta})$$

$$\begin{bmatrix} \dots & & & \\ \vdots & x_{i1} & x_{i2} & \dots & x_{ip} & \vdots \\ \dots & & & \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ \vdots \\ w_{p-1} \\ w_p \end{bmatrix} = \begin{bmatrix} \vec{\beta} \cdot \vec{x}_i \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \underset{p+1}{}$$

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

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

$$\hat{y} \ (\text{preds})$$

$$Y$$

Outline for today

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

- Introduction to applied linear algebra

- *Multiple* linear regression

- Analytic solution to multiple linear regression

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

take derivative

number (scalar)

$$J(\vec{w}) = \frac{1}{2}\left(\vec{y} \cdot \vec{y} - 2\vec{y} \cdot (X\vec{w}) + (X\vec{w}) \cdot (X\vec{w})\right)$$

$$\frac{dJ}{d\vec{w}} = -X^T\vec{y} + (X^TX)\vec{w} = 0$$

Solve for $\vec{w}$

$$\boxed{(X^TX)\vec{w} = X^T\vec{y}}$$

$(p+1) \times n \quad n \times (p+1)$

$$X\vec{w} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \sim p+1$$

$$\left(X^TX\right)^{-1}\left(X^TX\right)\vec{w} = \left(X^TX\right)^{-1}X^T\vec{y}$$

$$\boxed{\vec{w} = \underbrace{\left(X^TX\right)^{-1}}_{\text{Variance of } X}\underbrace{X^T\vec{y}}_{\text{Covariance of } X,y}}$$

Variance of X

Covariance of X, y

<u>Analytic solution to multiple linear regression</u>

(Keep this formula and its interpretation in mind!)

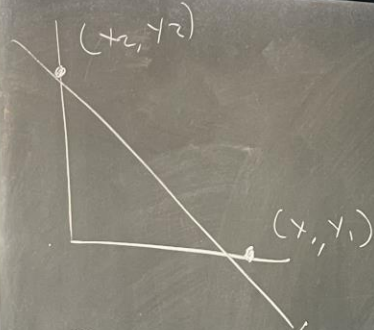① $AB = \begin{bmatrix} 4 & 1 \\ -8 & 8 \end{bmatrix}$ $BA = \begin{bmatrix} 6 & 4 \\ -1 & 6 \end{bmatrix}$

② $X = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\underbrace{}_{\text{fake}}$ $\underbrace{}_{x_1}$ $\underbrace{}_{y_1}$
ones x_2 y_2

(x_2, y_2)

(x_1, y_1)

③ $\vec{w} = (X^T X)^{-1} X^T \vec{y}$

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

Handout 5

Handout 5

(x_i, y_i)

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

$\left(\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\left\{ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right.$

$= \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
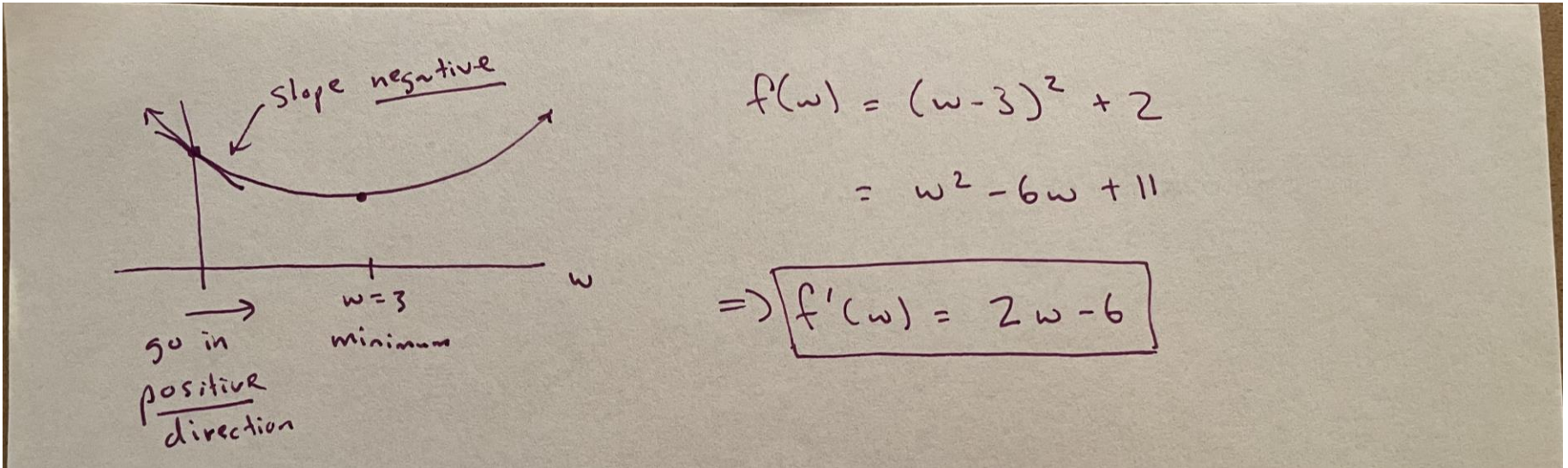
Same as before!

$\begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$f($

Begin gradient descent

Small example
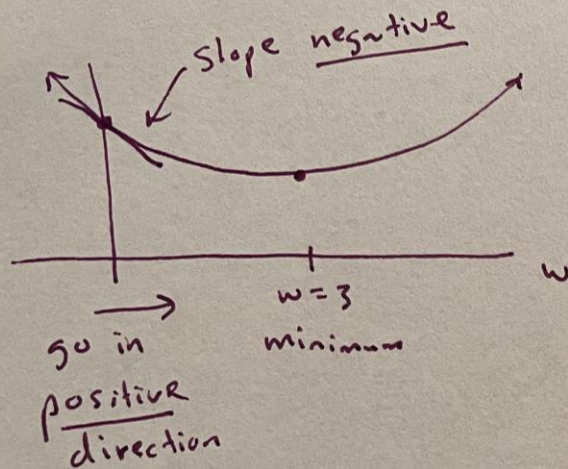
Goal: minimize the function $f(w) = w^2 - 6w + 11$

Slope negative

go in
positive
direction

$w = 3$
minimum

w

$f(w) = (w-3)^2 + 2$

$= w^2 - 6w + 11$

$\Rightarrow \boxed{f'(w) = 2w - 6}$

Small example

Goal: minimize the function $f(w) = w^2 - 6w + 11$

Slope negative

$f(w) = (w-3)^2 + 2$

$= w^2 - 6w + 11$

$\Rightarrow \boxed{f'(w) = 2w - 6}$

$w = 3$
minimum

go in
positive
direction

① $\quad w \leftarrow 0 - 0.1(2 \cdot 0 - 6)$

$w \leftarrow 0 + 0.6$

$\boxed{w \leftarrow 0.6}$

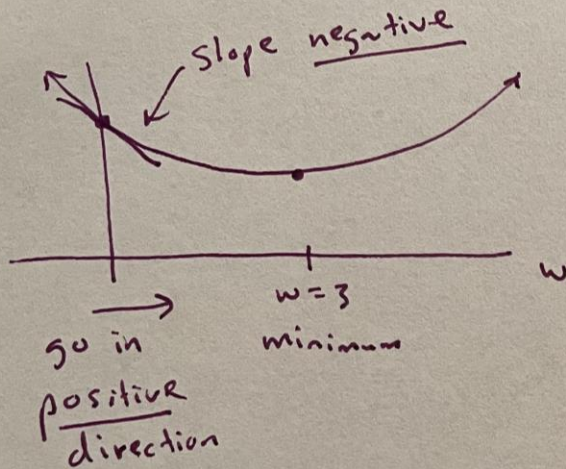② $\quad w \leftarrow 0.6 - 0.1(2 \cdot 0.6 - 6)$

$w \leftarrow 0.6 - 0.1(-4.8)$

$\boxed{w \leftarrow 1.08}$

$$w \leftarrow w - \alpha f'(w)$$

step size

Small example

Goal: minimize the function $f(w) = w^2 - 6w + 11$

slope negative

$w = 3$

go in positive direction

minimum

$f(w) = (w-3)^2 + 2$

$= w^2 - 6w + 11$

$\Rightarrow \boxed{f'(w) = 2w - 6}$

① $w \leftarrow 0 - 0.1(2 \cdot 0 - 6)$

$w \leftarrow 0 + 0.6$

$\boxed{w \leftarrow 0.6}$

② $w \leftarrow 0.6 - 0.1(2 \cdot 0.6 - 6)$

$w \leftarrow 0.6 - 0.1(-4.8)$

$\boxed{w \leftarrow 1.08}$

Stop when:

$|f(w^t) - f(w^{t-1})| < \varepsilon$, $\varepsilon = 1 \times 10^{-8}$

(for example)