CS 260: Foundations of Data Science

Prof. Sara Mathieson

Spring 2025

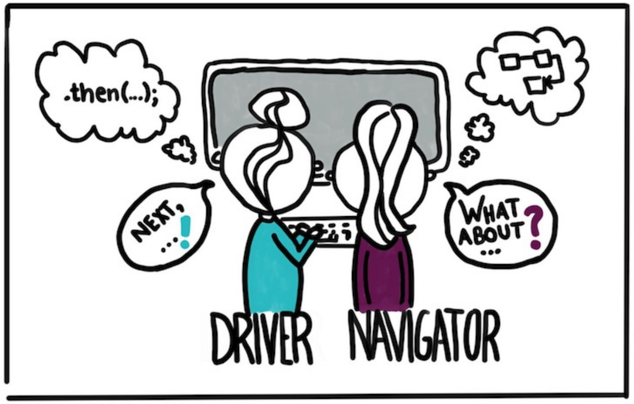
HAVERFORD
COLLEGE

Admin

- Lab 2 posted, due Wednesday
- TA hours tonight! 8-10pm in H110

Pair Programming

- One person is **driver** (at the keyboard)
- One person is **navigator**
- Switch every 30 min!
- Always be working on the assignment together
- No “divide and conquer”
- Make sure to push frequently and pull if there have been any changes



https://dev.to/documatic/pair-programming-best-practices-and-tools-154j

Outline

- Why are models useful? (recap)
- Linear models
- Fitting a linear model (one feature)
- Model complexity and evaluation

Outline

- Why are models useful? (recap)
- Linear models
- Fitting a linear model (one feature)
- Model complexity and evaluation

Why are models useful?

- Understand/explain/interpret the phenomenon

- Predict outcomes for future examples

Outline

- Why are models useful? (recap)
- Linear models
- Fitting a linear model (one feature)
- Model complexity and evaluation

Goals of fitting a linear model

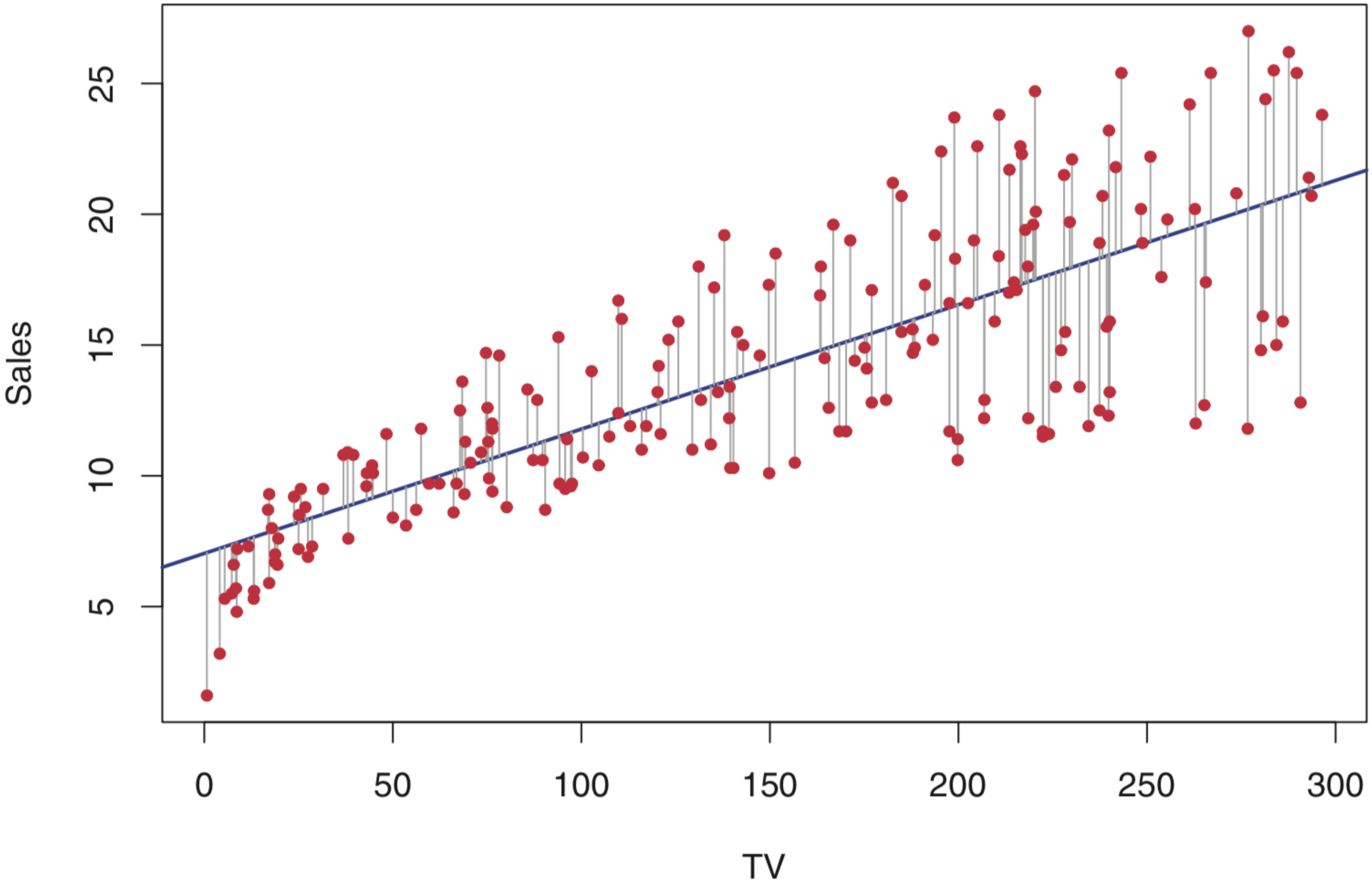
1) Which of the features/explanatory variables/predictors (x) are associated with the response variable (y)?

2) What is the relationship between x and y?

3) Is a linear model enough?

4) Can we predict y given a new x?

Example: predict sales from TV advertising budget



ISL: Figure 3.1

Maybe a linear model is not enough

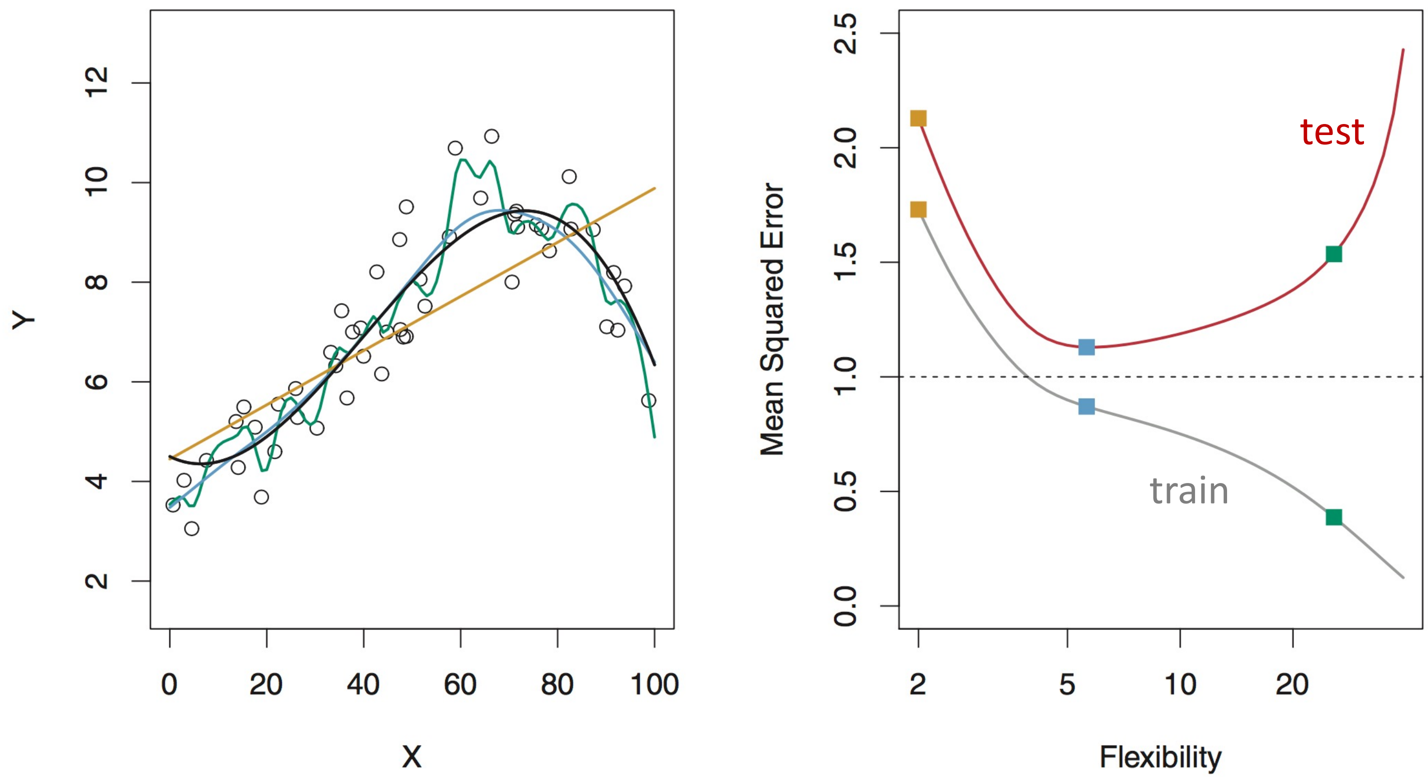
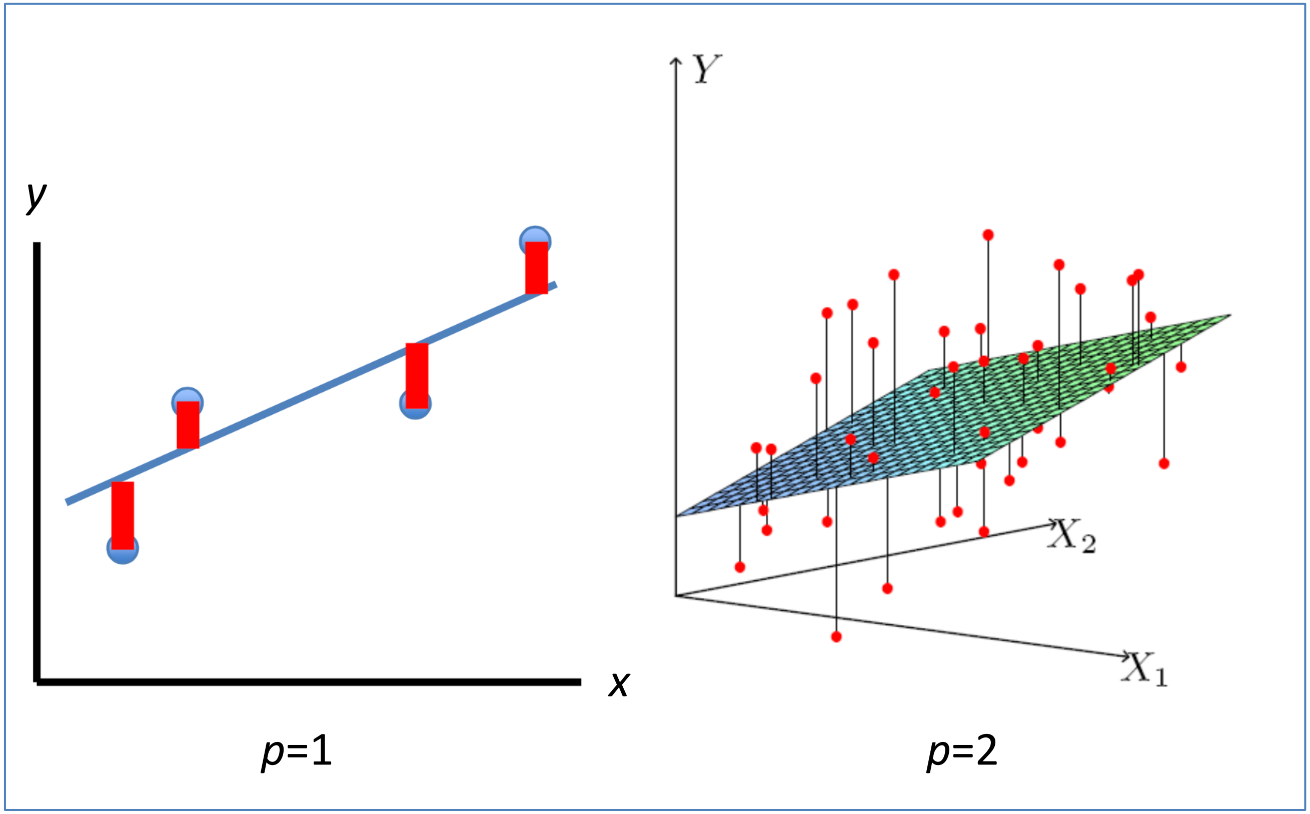


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed line). Squares represent the training and test MSEs for the three fits shown in the left-hand panel.*

Linear model with 1 or 2 features



Slide: modified from Jessica Wu
Original: Eric Eaton

Linear Regression

- Output (*y*) is continuous, not a discrete label

- Learned model: *linear function* mapping input to output (a *weight* for each feature + *bias*)

- Goal: minimize the *RSS* (residual sum of squared errors) or *SSE* (sum of squared errors)

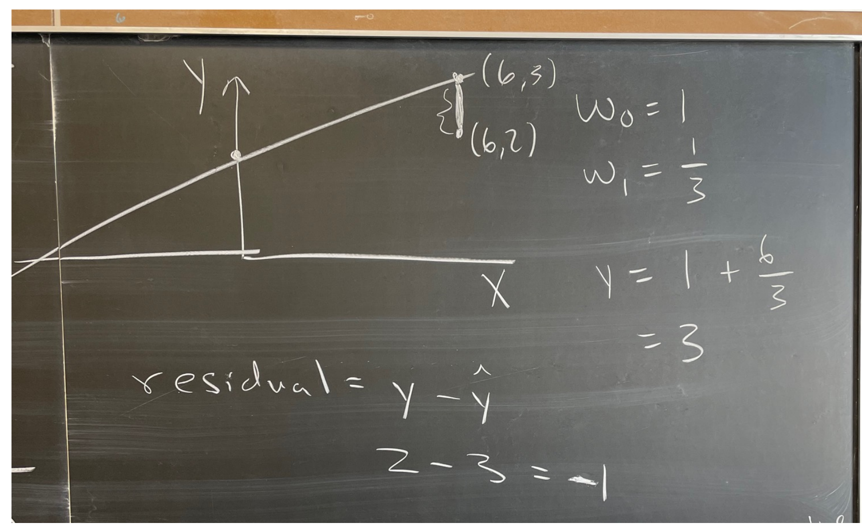
Outline

- Why are models useful? (recap)
- Linear models
- Fitting a linear model (one feature)
- Model complexity and evaluation

Mini-quiz (discuss with a partner)

Say we have the linear model: $y = 1 + x/3$

1) Sketch a graph of this line
2) What is the slope? What is the y-intercept?
3) What parameters do these correspond to in our linear model?
4) If we have a point $(x_1,y_1) = (6,2)$, what is the residual?

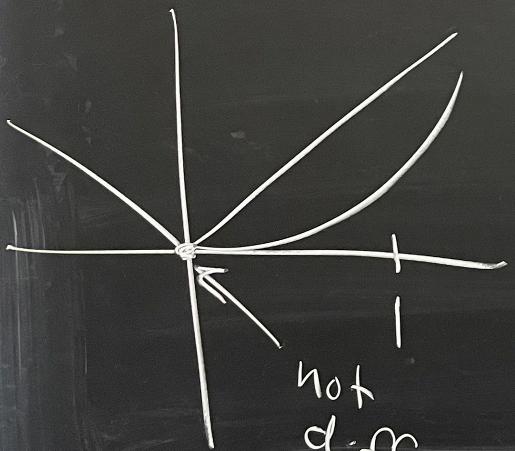


model $h_{\vec{w}}(x) = \underbrace{w_0 + w_1 x} = \hat{y}$ (pred)

GOAL minimize:

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

$\cancel{|y_i - \hat{y}_i|}$



cost function (loss)

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

makes math nice

$$\underbrace{\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1} \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\text{fixed!}}$$

want

(a) $\dfrac{\partial J}{\partial w_0} = 0$

(b) $\dfrac{\partial J}{\partial w_1} = 0$

(a

$\bar{X} =$

$\bar{y} =$

(a)

$$\frac{\partial J}{\partial w_0} = -\sum_{i=1}^{n} (y_i - w_0 - w_1 x_i) = 0$$

$$-\sum y_i + n w_0 + w_1 \sum x_i = 0$$

$$w_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - w_1 \frac{1}{n}\sum_{i=1}^{n} x_i$$

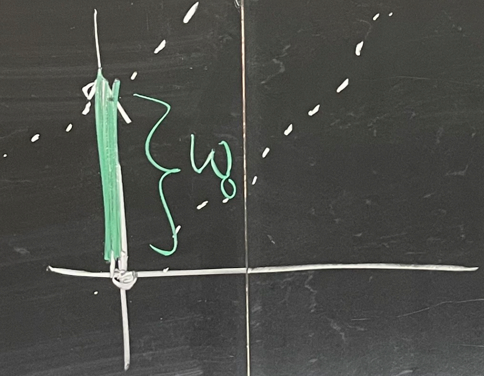
$\bar{x}$ = avg of all x_i's

$\bar{y}$ = avg of all y_i's

$$\boxed{\hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}}$$

(b)

$$\frac{\partial J}{\partial w_1} = 0$$



$$+\left(\frac{1}{n}\sum y_i\right) - \left(\bar{y}\right)$$

(b) $\frac{\partial J}{\partial w_1}$

$\hat{w}_1$

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

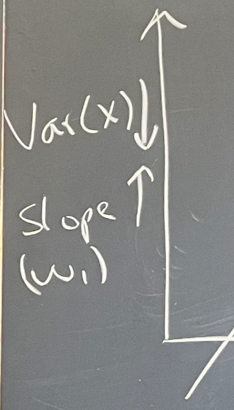
$$= -\sum_{i=1}^{n} (y_i x_i - w_0 x_i - w_1 x_i^2) = 0$$

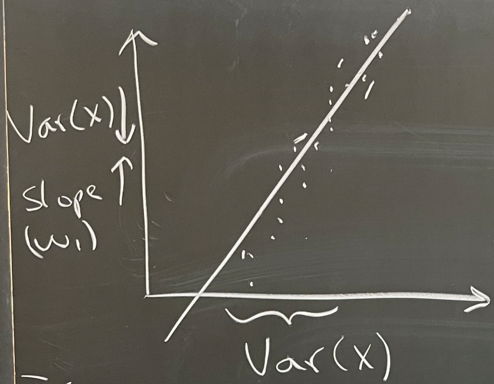
$$= -\sum_{i=1}^{n} (y_i x_i - \bar{y} x_i + w_1 \bar{x} x_i - w_1 x_i^2) = 0$$

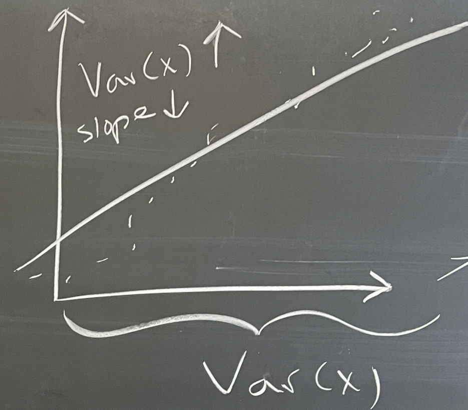
solve for w_1

$$\boxed{\hat{w}_1 = \frac{\sum_{i=1}^{n} (y_i x_i - \bar{y} x_i)}{\sum_{i=1}^{n} (x_i^2 - \bar{x} x_i)}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} =$$

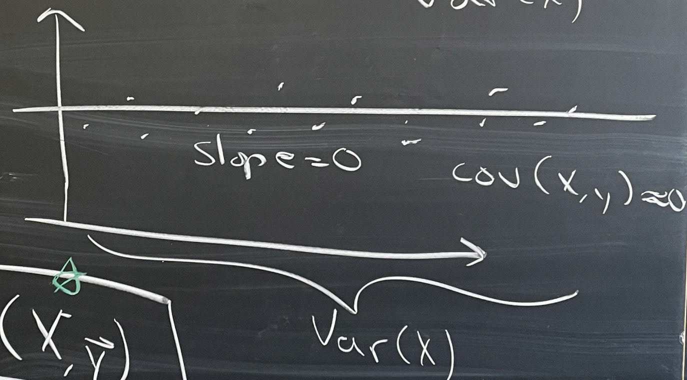
Skip!
(add 0)





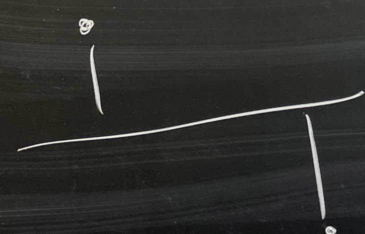


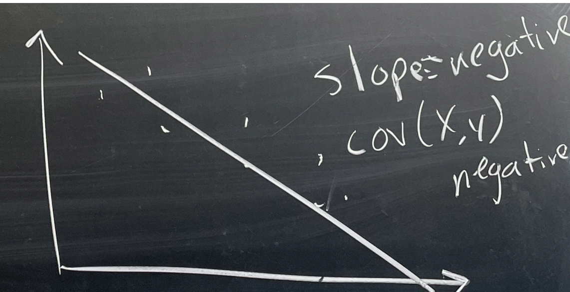
$x_i^2) = 0$



$$\frac{(y_i - \bar{y})}{)^2} = \boxed{\frac{Cov(X, \vec{Y})}{Var(X)}}$$

$) \approx 0$





Handout 4

- Work with your partner for Lab 2
- Exchange contact information!

Handout 4

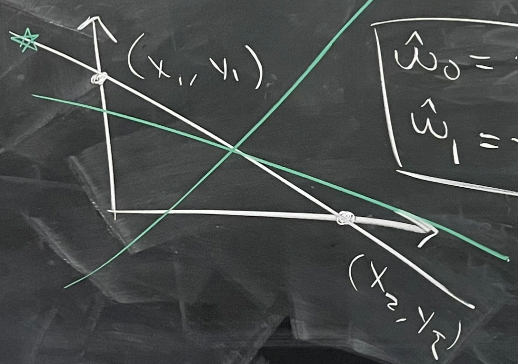
② $p = 1$

$\# \text{params} = 2 \quad (w_0 + w_1)$

③ <u>cost / loss</u>

$$J(w_0, w_1) = \frac{1}{2} \sum_{i=1}^{n} \left(y_i - \underbrace{w_0 + w_1 x_i}_{\hat{y}_i}\right)^2$$

④



$$\boxed{\begin{aligned} \hat{w}_0 &= 1 \\ \hat{w}_1 &= -1 \end{aligned}}$$

⑤

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

$$\boxed{\hat{w}_1 = -1}$$

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

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

$$\hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}$$

$\bar{x}$ = avg

$$\frac{1}{2} + 1 \cdot \frac{1}{2}$$

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

Outline

- Why are models useful? (recap)
- Linear models
- Fitting a linear model (one feature)
- Model complexity and evaluation

$0-\frac{1}{2})$

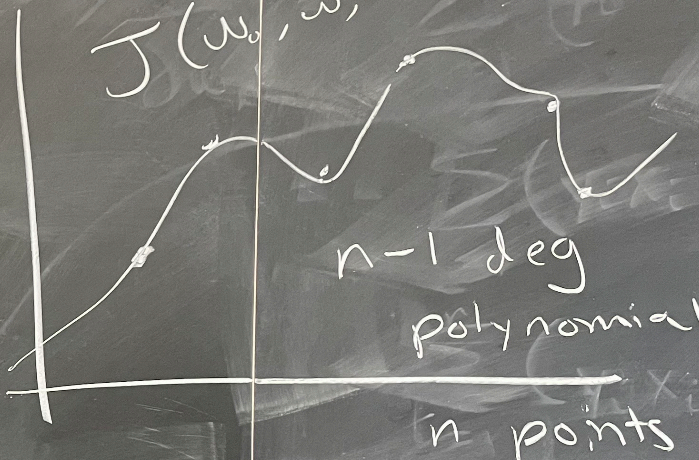
Model complexity

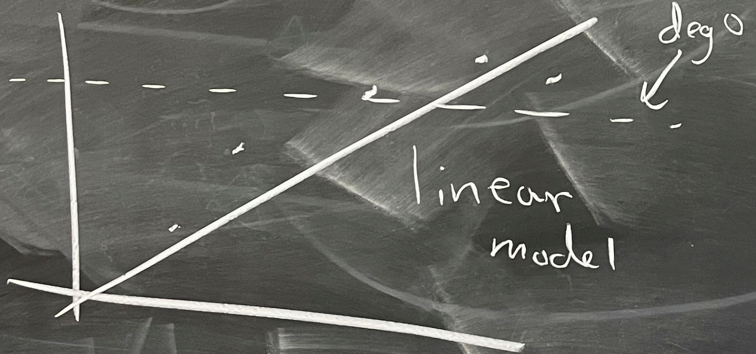
$\approx$ # parameters

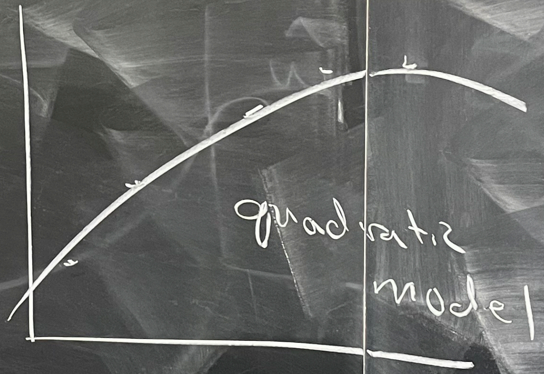
$\hat{y} = w_0 + w_1 x + w_2 x^2$

3 parameters

$J(w_0, w_1, \dots w_{n-1}) = 0$

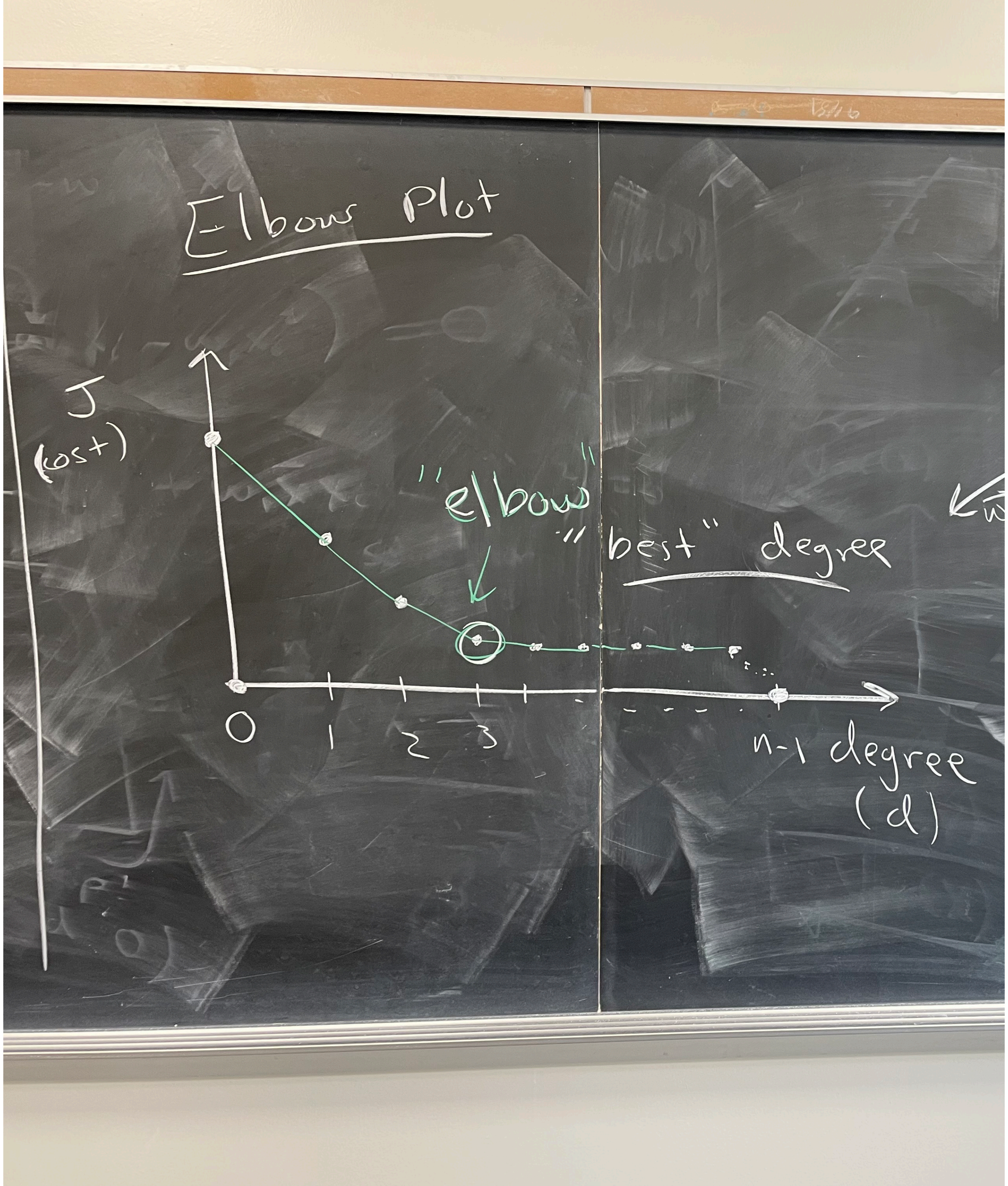






J

(cost


Elbow Plot
J
(cost)
"elbow"
"best" degree
0
1
2
3
n-1 degree
(d)

Intro to Vectors and Matrices

Vectors

- Vector magnitude

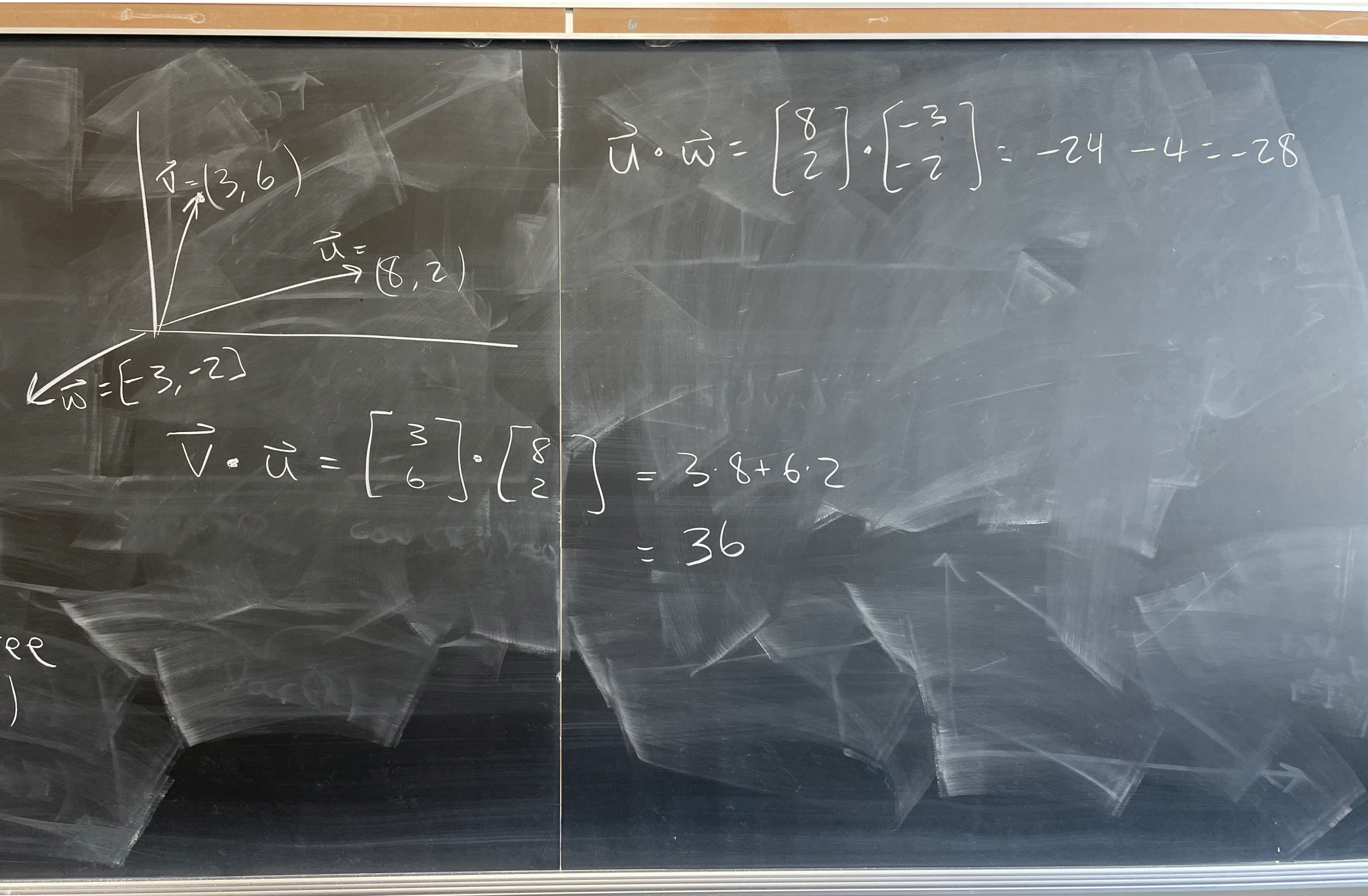
$$\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \text{then} \quad |\mathbf{v}| = \sqrt{x^2 + y^2}.$$

- Different ways to write a vector

$$\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix}^{\mathrm{T}}$$

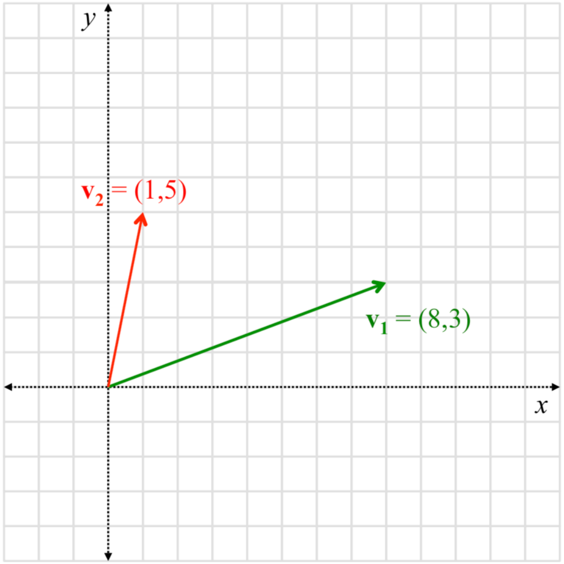
- Vector dot product

Vector Dot Product



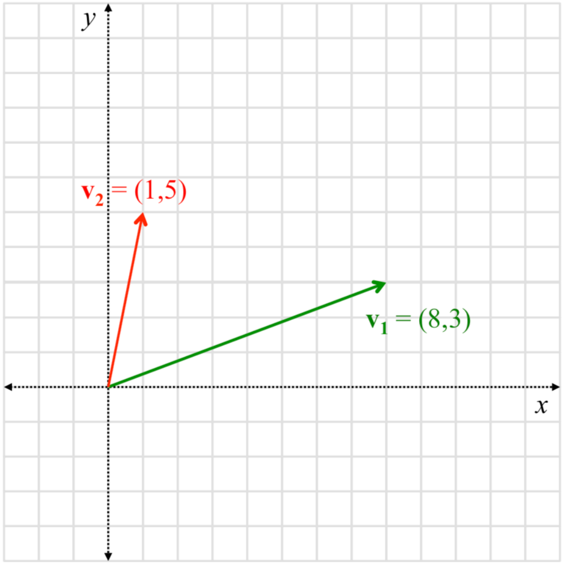
Vector Addition

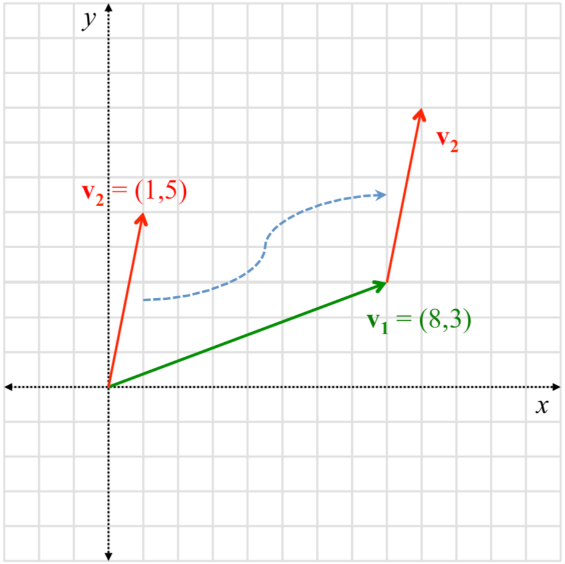
$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 8+1 \\ 3+5 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$



Vector Addition

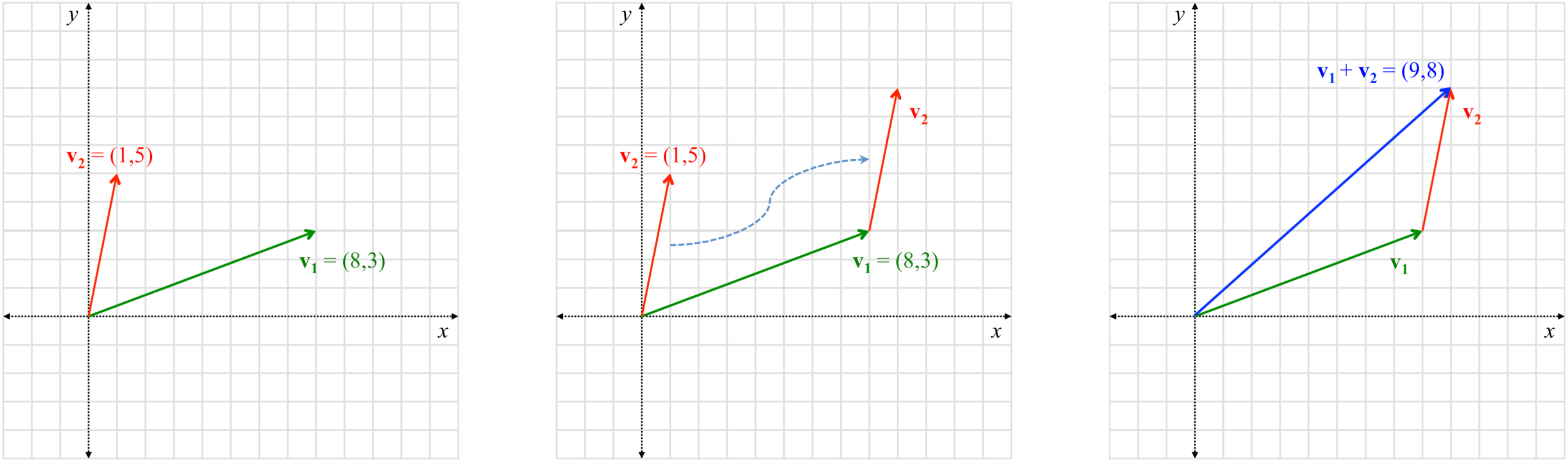
$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 8+1 \\ 3+5 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$





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!)

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