CS 260: Foundations of Data Science

Prof. Thao Nguyen

Fall 2025

HAVERFORD COLLEGE

Materials by Sara Mathieson

Admin

- Sit somewhere new
- Lab deadline - 11:59pm

Outline for today

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

Outline for today

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

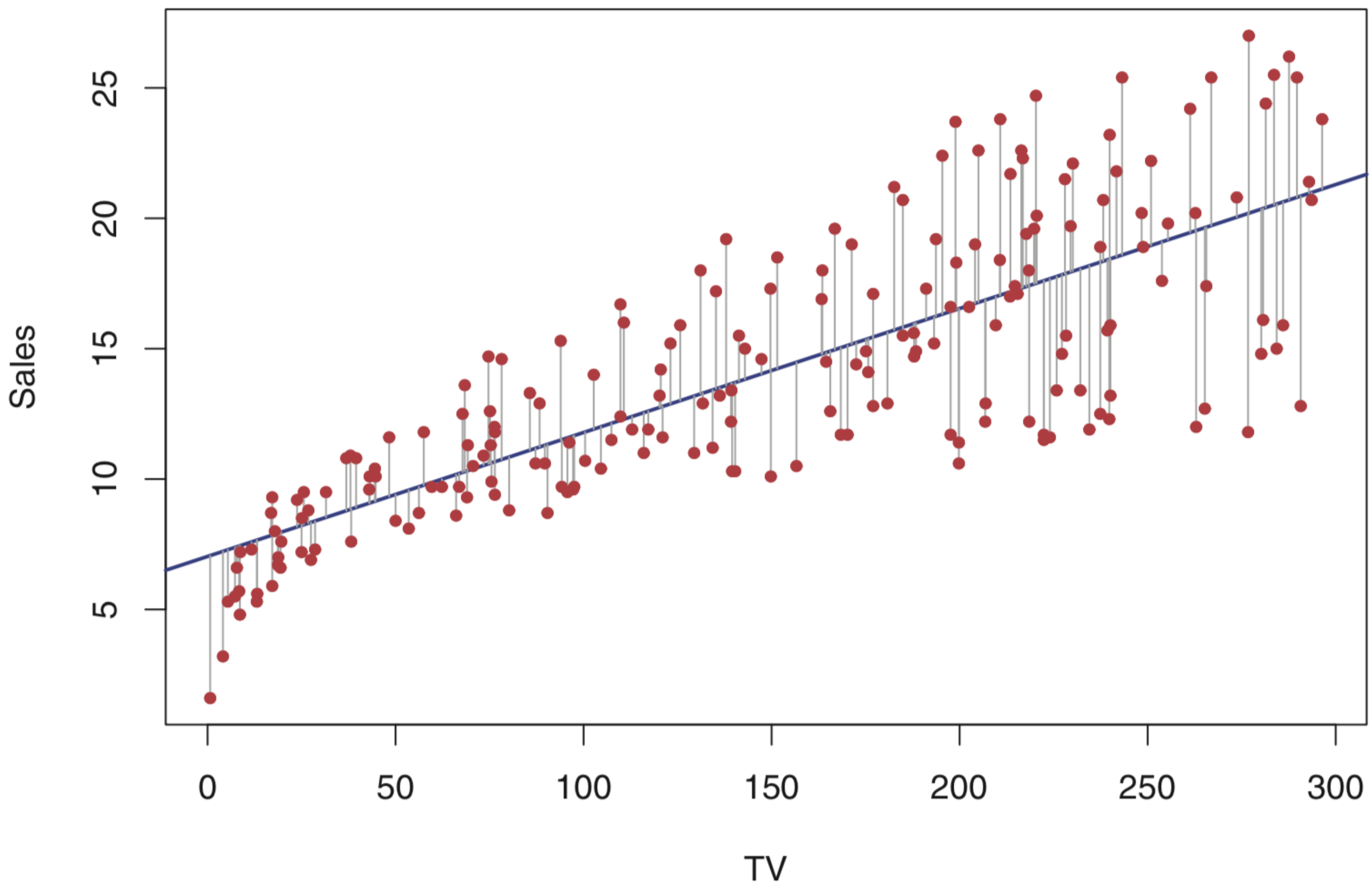
Why are models useful?

- Understand/explain/interpret the phenomenon
- Predict outcomes for future examples

Outline for today

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

Example: predict sales from TV advertising budget



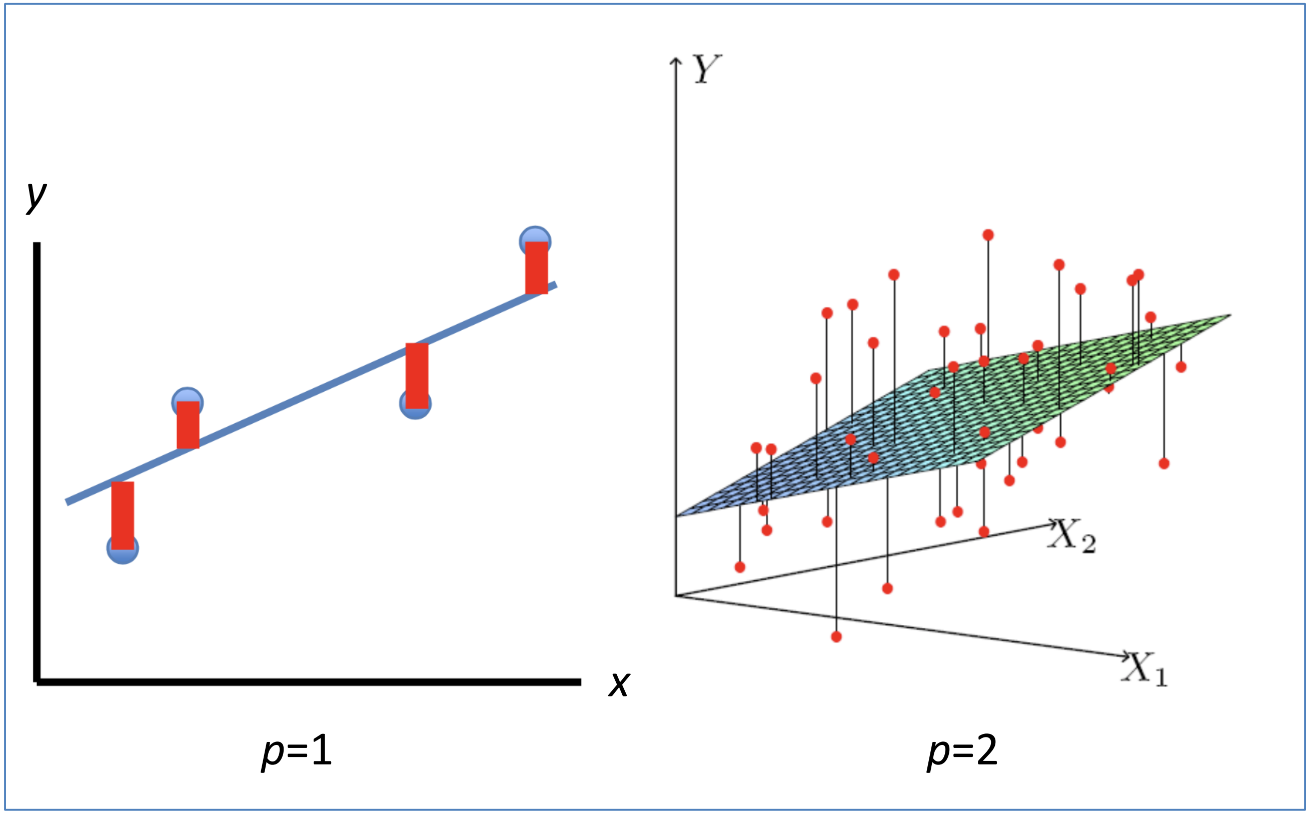
ISL: Figure 3.1

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 $(x1,y1) = (6,2)$, what is the residual?

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 squares) or *SSE* (sum of squared errors)

Outline for today

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

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

GOAL minimize:

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

[redacted]

cost function (loss)

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

makes math nice

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

fixed!

want

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

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

(a

$\bar{X} =$

$\bar{y} =$

Fitting a linear model

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

$$\leftrightarrow\ -\sum_{i=1}^{n} y_i + n w_0 + \sum_{i=1}^{n} w_1 x_i = 0$$

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

$$\leftrightarrow\ \boxed{\widehat{w_0} = \bar{y} - w_1 \bar{x}}$$

$\bar{x}$: avg of all x_i's
$\bar{y}$: avg of all y_i's

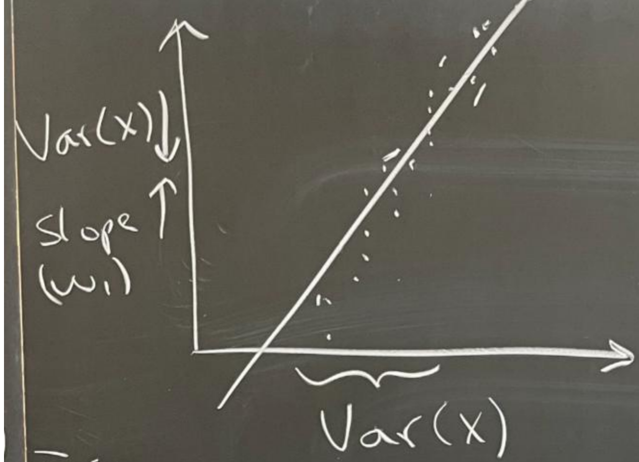
Fitting a linear model

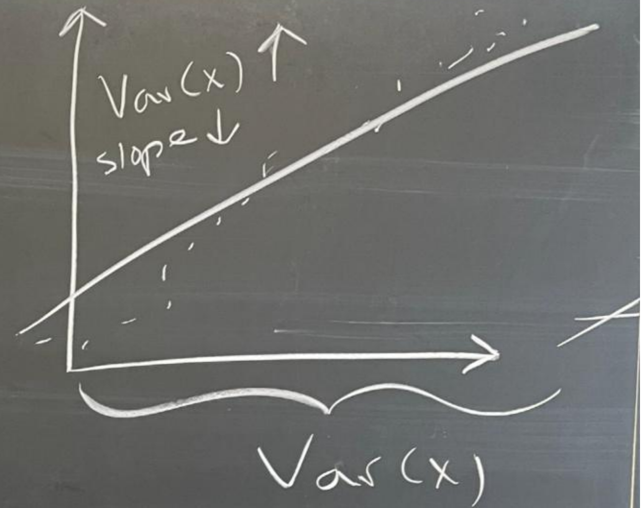
b) $\frac{\partial J}{\partial w_1} = -\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i)x_i = 0$

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

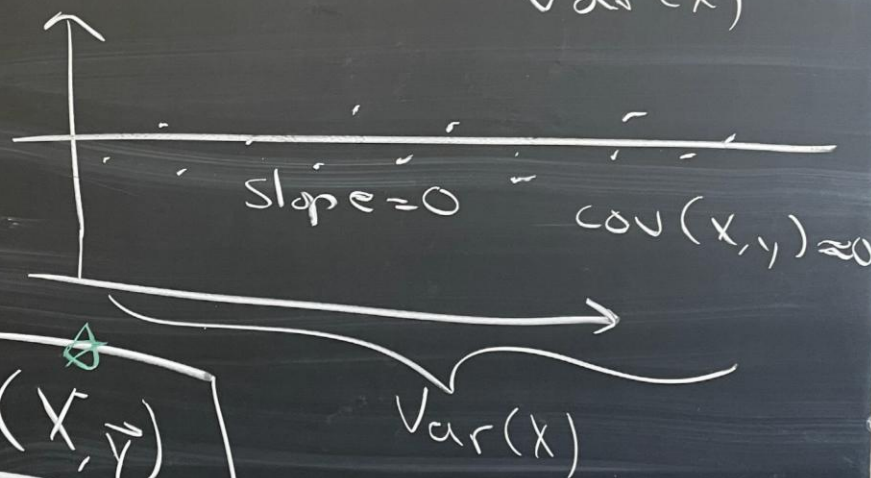
$$\leftrightarrow w_1 = \frac{\sum_{i=1}^{n}(y_i x_i - \bar{y}x_i)}{\sum_{i=1}^{n}(x_i^2 - \bar{x}x_i)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\leftrightarrow \boxed{\widehat{w_1} = \frac{Cov(X, \vec{y})}{Var(X)}}$$



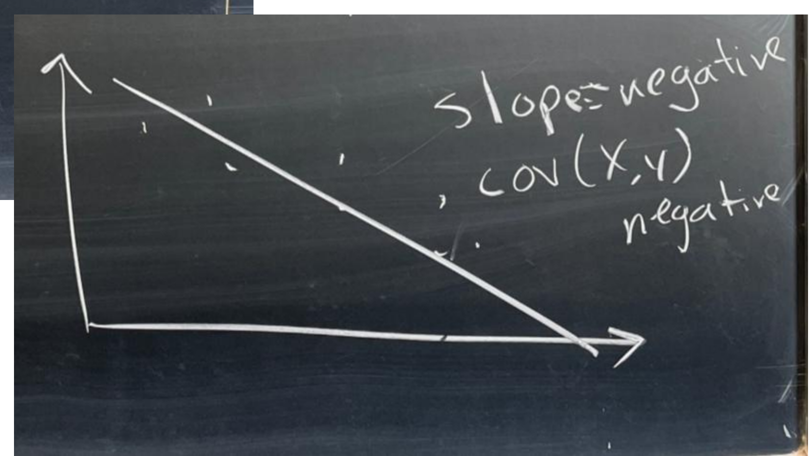


$= 0$



$-\bar{y})$

$$= \frac{Cov(X, \vec{y})}{Var(X)}$$



Handout 4

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?

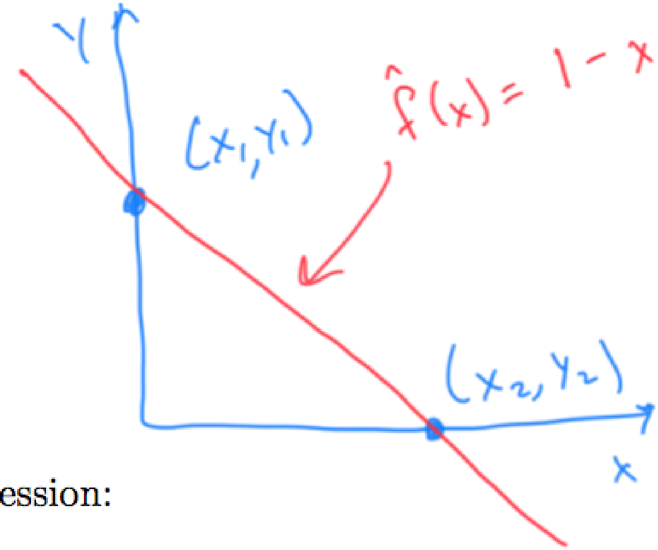
3) Can we (accurately) predict y given a new x?

4) Is a linear model enough?

Handout 4

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).

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

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

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

Outline for today

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

Maybe a linear model is not enough

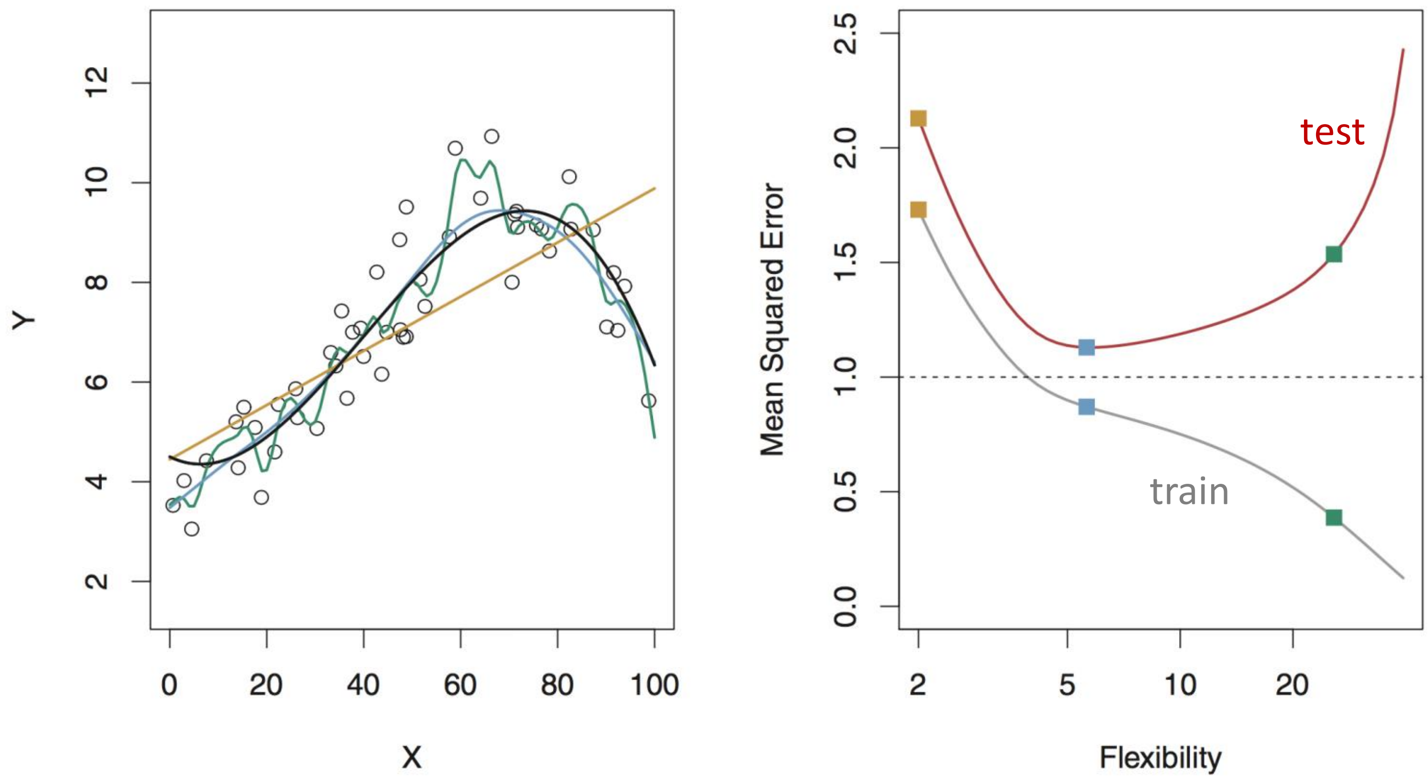
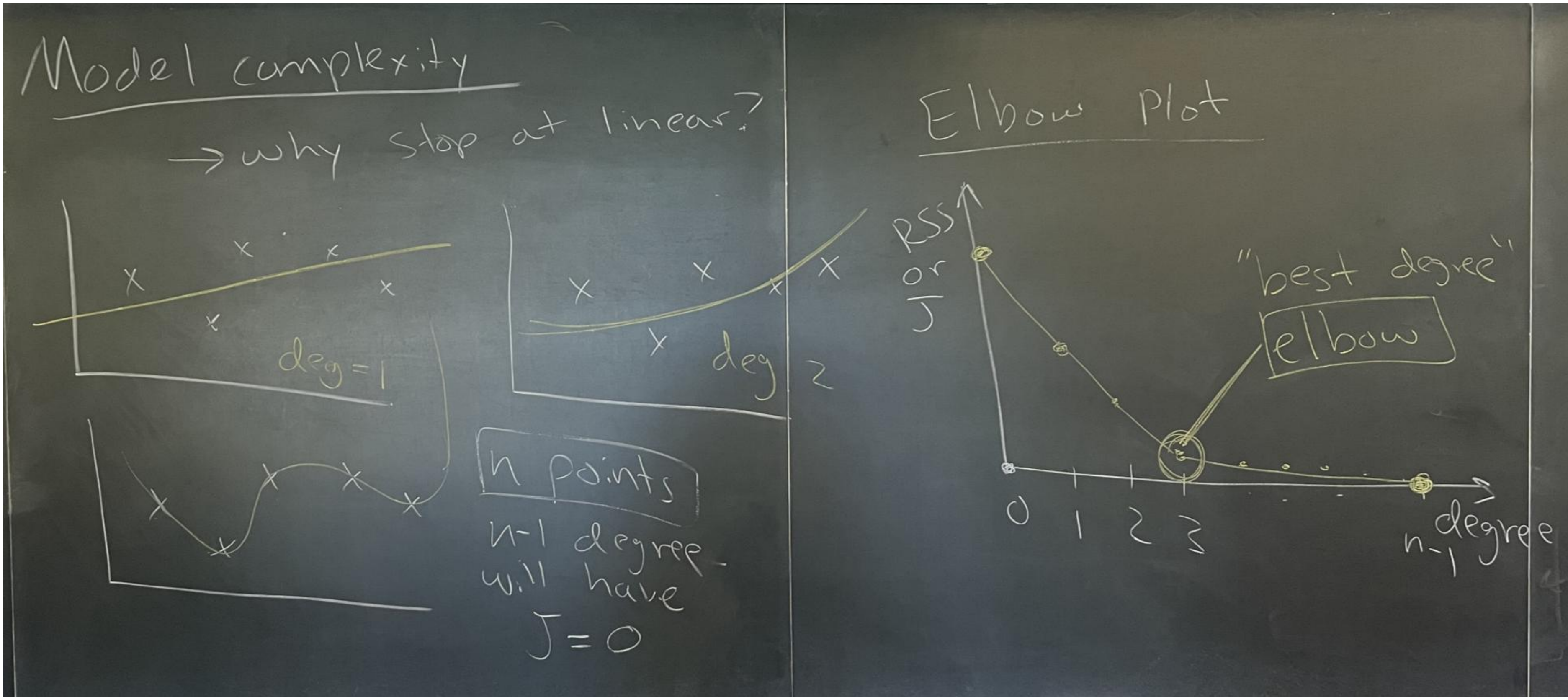


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed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

ISL, Fig 2.9


Model complexity
→ why stop at linear?
deg = 1
deg 2
n points
n-1 degree will have
J = 0
Elbow Plot
RSS or J
"best degree"
elbow
0
1
2
3
degree
n-1