CS 260: Foundations of Data Science

Prof. Sara Mathieson

Fall 2021

HAVERFORD COLLEGE

Admin

- Lab 2 posted, due Tuesday
- Notetaker: Sharon

TA Hour Schedule

Monday 7-8:30pm: Trang (BMC campus)
Tuesday 6:30-8:30pm: Nasa (H110)
Wednesday 4-6pm: Yuxuan (H204)

Outline for September 9

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

Outline for September 9

- 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

What are the most important features?

X

Color	Shape	Size
red	square	big
blue	square	big
red	circle	small
blue	square	small
red	circle	big

Y

Likes toy?
+
+
-
-
+

Example by: Ameet Soni

What are the most important features?

X			Y
Color	**Shape**	**Size**	**Likes toy?**
red	square	big	+
blue	square	big	+
red	circle	**big**	-
blue	square	**big**	-
red	circle	big	+

Example by: Ameet Soni

Outline for September 9

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

Goals of fitting a linear model

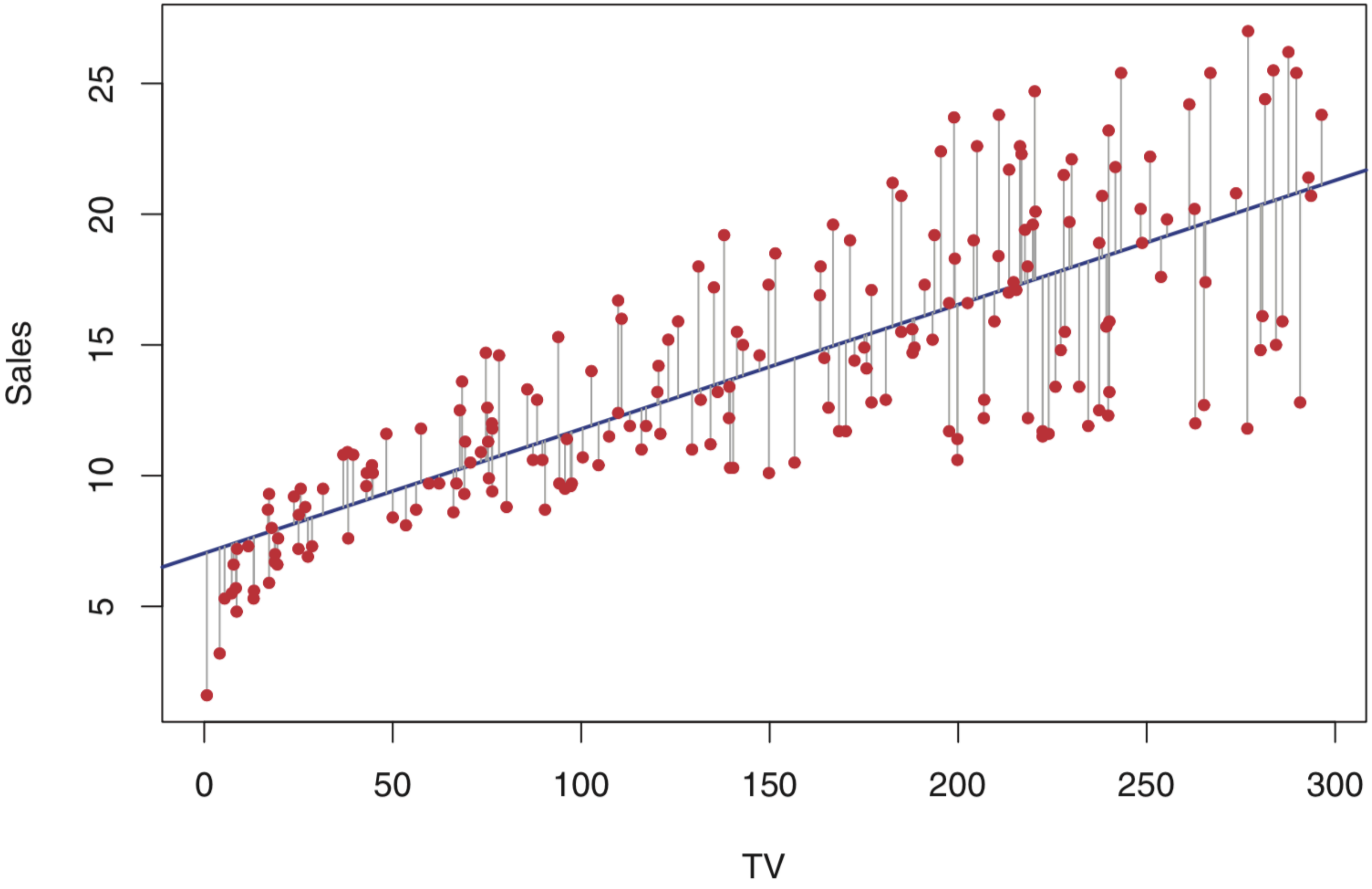
1) Which of the features/explanatory variables/predictors (x) are associated with the response variable (y)?

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

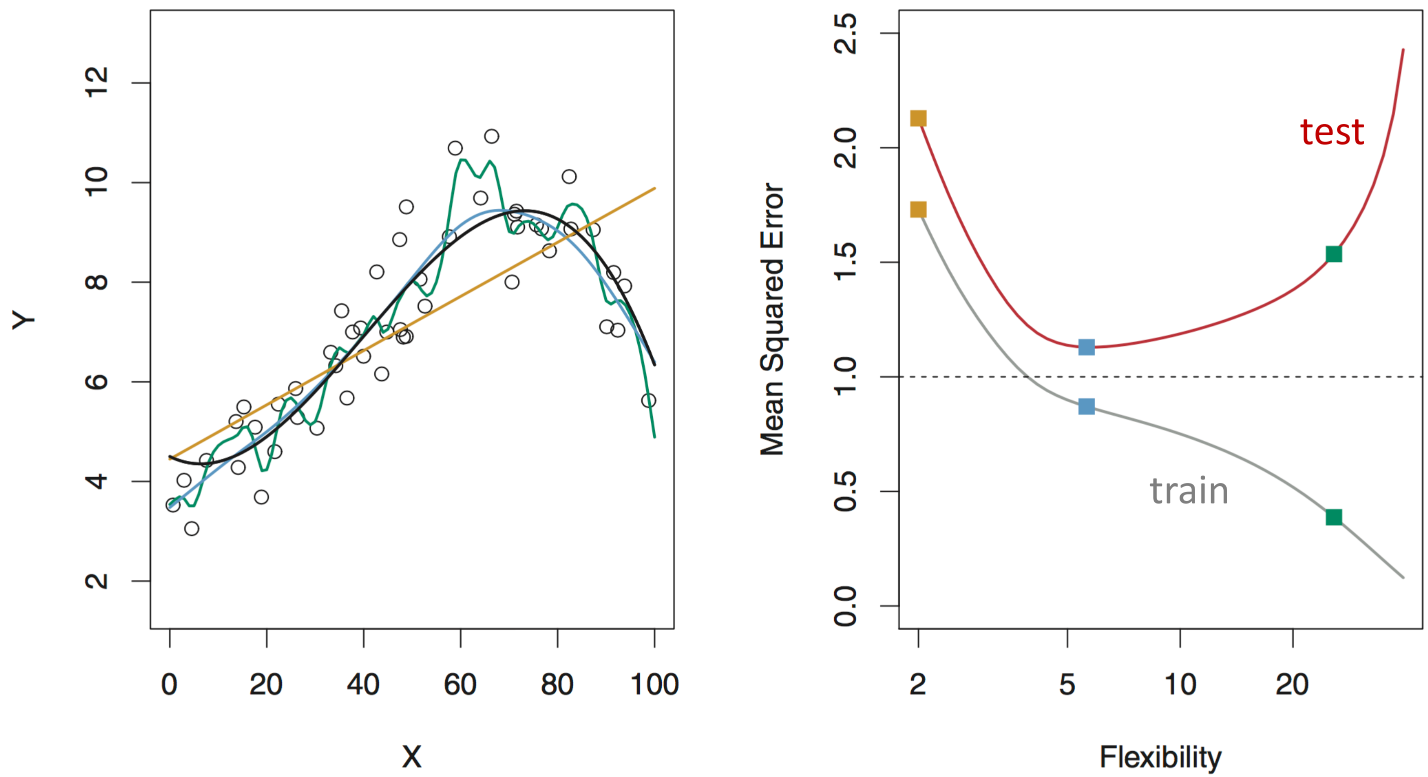
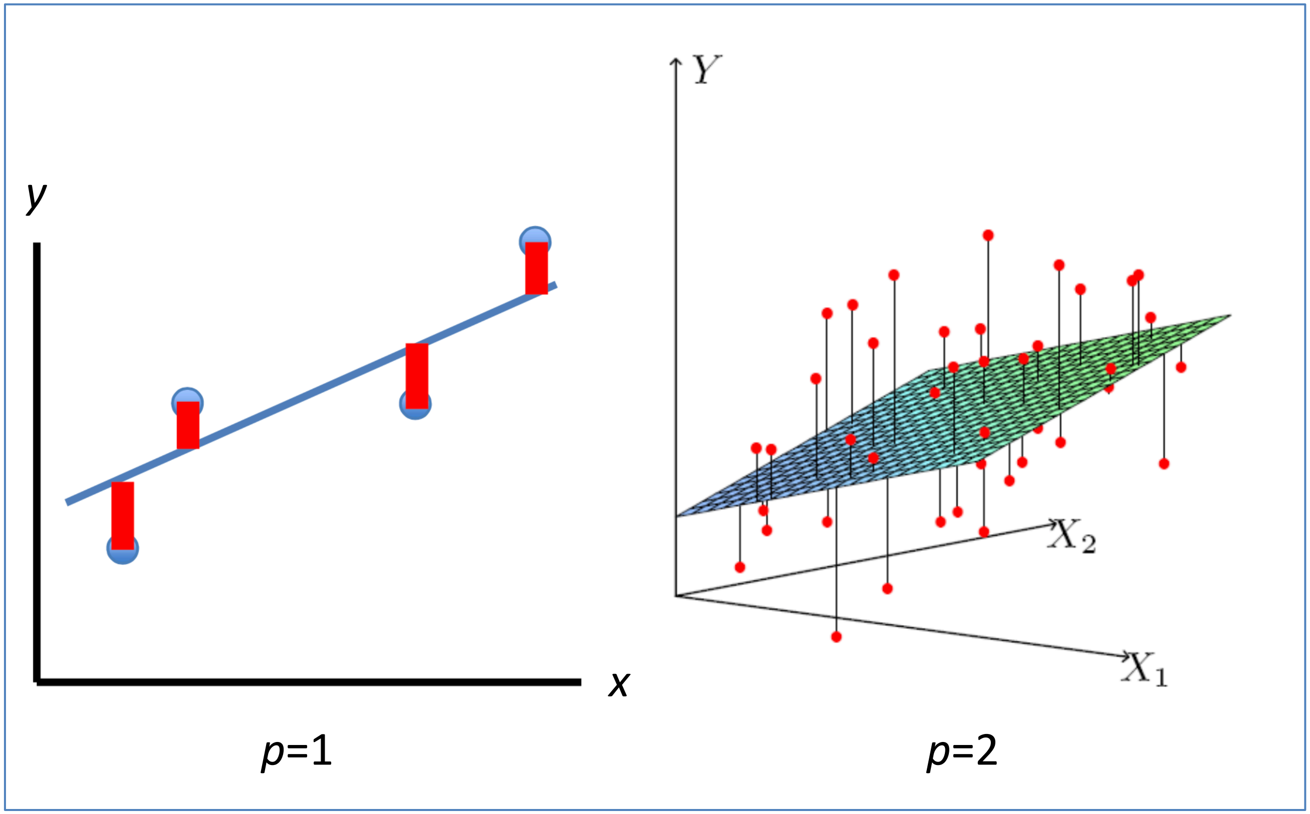


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed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

ISL, Fig 2.9

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 for September 9

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

PLEASE LEAVE COMPUTERS ON
do not erase

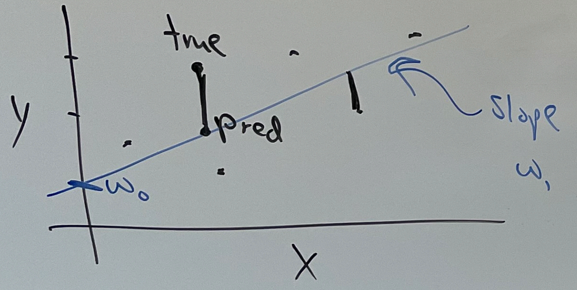
Linear Model

* feature: x
* response: y

Model: $h_{\vec{w}}(x) = \overbrace{\underbrace{w_0 + w_1}_{\text{model params}} x}^{\hat{y}}$

hypothesis.

$\vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$



"fit" the model: find $\hat{w}_0, \hat{w}_1$

residual: $y_i - \hat{y}_i$ (difference between true + predicted)

overall: minimize $\boxed{RSS}$: residual sum of squares

cost/loss function — minimize!

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

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

derivative wrt w_0

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

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

$\bar{x}$ = mean of x_i's

$\bar{y}$ = mean of y_i's

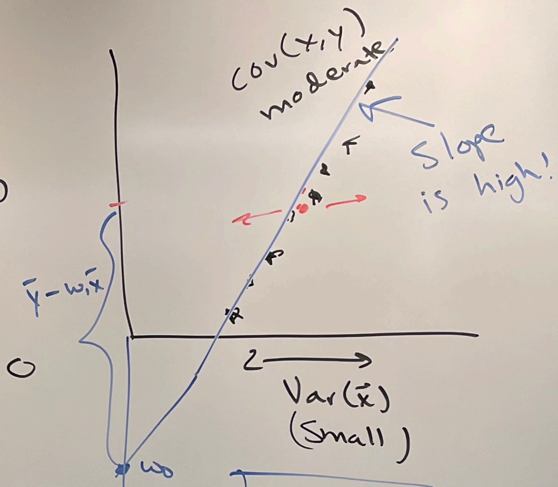
(b) $\boxed{w_1}$ $\quad \frac{\partial J}{\partial w_1} = -\sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)(x_i) = 0$

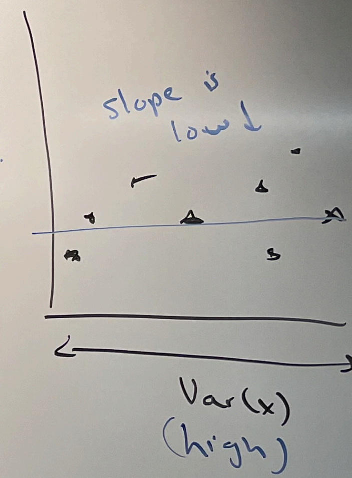
$-w_0 \cdot x_i$

$$-\sum \left(y_i x_i - \bar{y} x_i + w_1 \bar{x} x_i - w_1 x_i^2 \right) = 0$$

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

MANY STEPS





PLEASE LEAVE COMPUTERS ON

do not erase

Handout 4

② $p = 1$ ("simple")

\# model params = 2

③ $J(w_0, w_1)$ "cost" "loss"

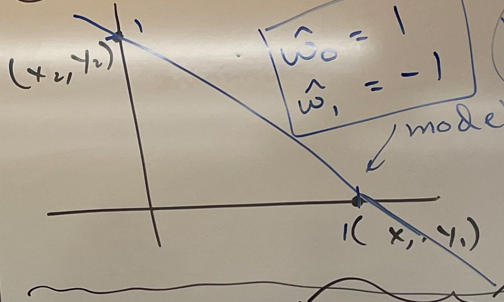
$= \frac{1}{2}\sum \ldots$

$= \boxed{\frac{1}{2}\,RSS}$

④ "x" $\rightarrow$ $\vec{x} = \begin{array}{c} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{array}\begin{array}{l} x_1 \\ x_2 \end{array}$ $\quad \vec{y} = \begin{array}{c} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{array}\begin{array}{l} y_1 \\ y_2 \end{array}$

$(x_1, y_1) = (1, 0)$

$(x_2, y_2) = (0, 1)$



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

whole model!!

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

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

$$\hat{w}_1 = \frac{\overbrace{(1-\frac{1}{2})(0-\frac{1}{2})}^{(x_1, y_1)} + \overbrace{(0-\frac{1}{2})(1-\frac{1}{2})}^{(x_2, y_2)}}{(1-\frac{1}{2})^2 + (0-\frac{1}{2})^2}$$

$$= \frac{-\frac{1}{4} - \frac{1}{4}}{\frac{1}{4} + \frac{1}{4}} = \boxed{-1}$$ ☆

Outline for September 9

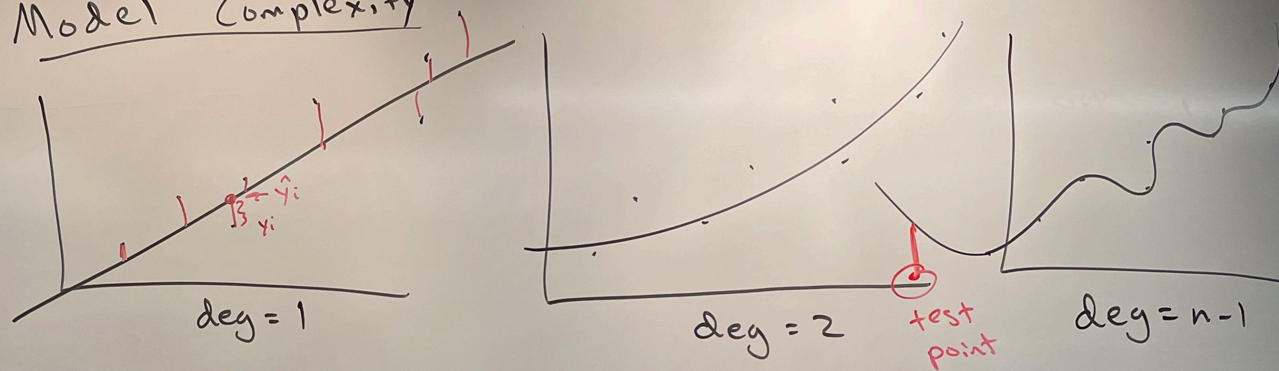
- Why are models useful? (recap)
- Linear models
- Fitting a linear model (one feature)
- Model complexity and evaluation

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

$$= \frac{1}{2} - (-1)\frac{1}{2}$$

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

Model Complexity

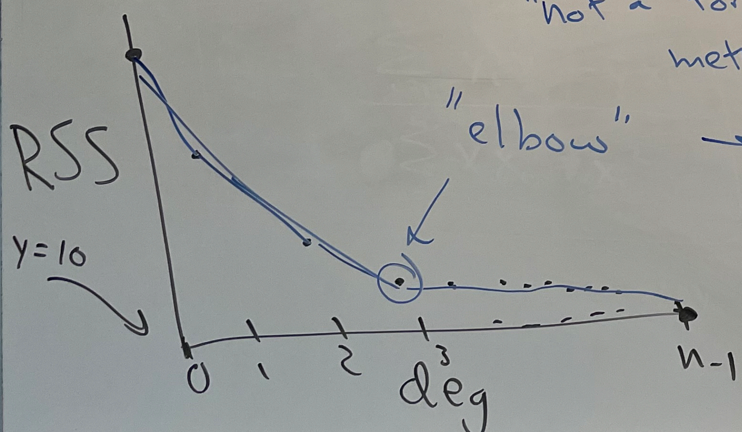


RSS = 0

great!

Elbow plot

"not a formal method"



→ $\boxed{\text{choice = deg 3}}$