CS 260 – Foundations of Data Science

Lecture 3 – Linear models

09/08/2026

Adam Poliak

Announcements

HW1 due Tuesday night (tonight)

Complete pre-course survey

Office Hours:

Thursdays 2-4pm, Park 200D

Outline

Why are models useful?

Linear models

Fitting a linear model (one feature)

Model complexity and evaluation

Why are models useful?

Understand/explain/interpret the phenomena

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

Outline

Why are models useful?

Linear models

Fitting a linear model (one feature)

Model complexity and evaluation

Linear Models

Features: $\vec{x}$ (x if p=1)

Output: $y \in \mathbb{R}$

b m

Model: $h_{\vec{w}}(x) = w_0 + w_1 x = \hat{y}$ prediction

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

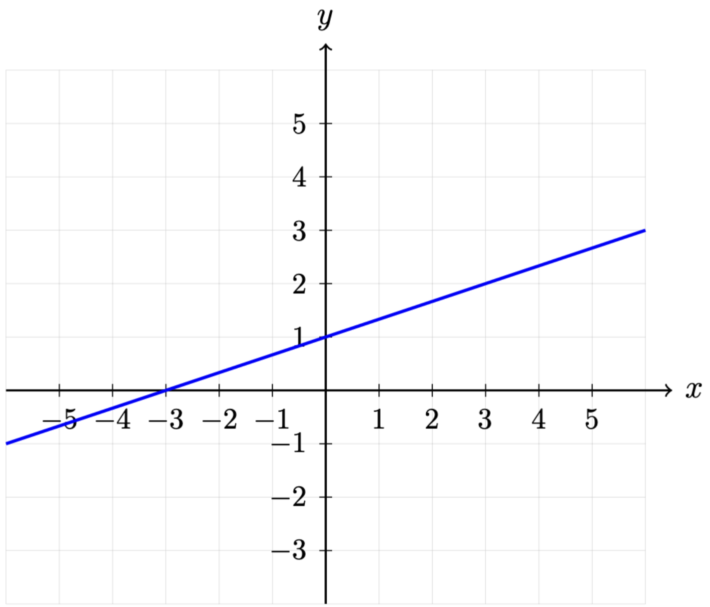
Goals:

Describe **linear dependence**

Predict output given new data

Handout

$$\textbf{slope} = \frac{1}{3}, \qquad y\textbf{-intercept} = 1.$$



Answer: The line crosses the y-axis at $(0, 1)$ and has slope $\frac{1}{3}$. For example, it also passes through $(3, 2)$ and $(6, 3)$.

Answer: Writing a linear model as

$$y = \beta_0 + \beta_1 x,$$

the y-intercept is β_0 and the slope is β_1. Therefore,

$$\beta_0 = 1, \qquad \beta_1 = \frac{1}{3}.$$

4. Suppose we have the point

$$(x_1, y_1) = (6, 2).$$

What is the residual for this point?

Answer: First, the model predicts

$$\hat{y}_1 = 1 + \frac{6}{3} = 3.$$

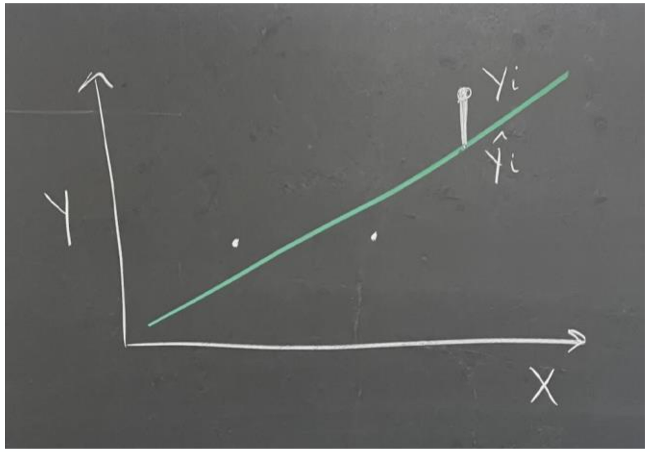
The residual is the observed value minus the predicted value:

$$e_1 = y_1 - \hat{y}_1 = 2 - 3 = \boxed{-1}.$$

Linear Models

How good is our model?

Residuals: $\underbrace{y_i}_{\text{truth}} - \underbrace{\hat{y}_i}_{\text{prediction}}$ } one example



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

RSS: Residual sum of squares
SSE: Sum of squared errors

Goal: Find a model that minimizes SSE

Goals of fitting a linear model

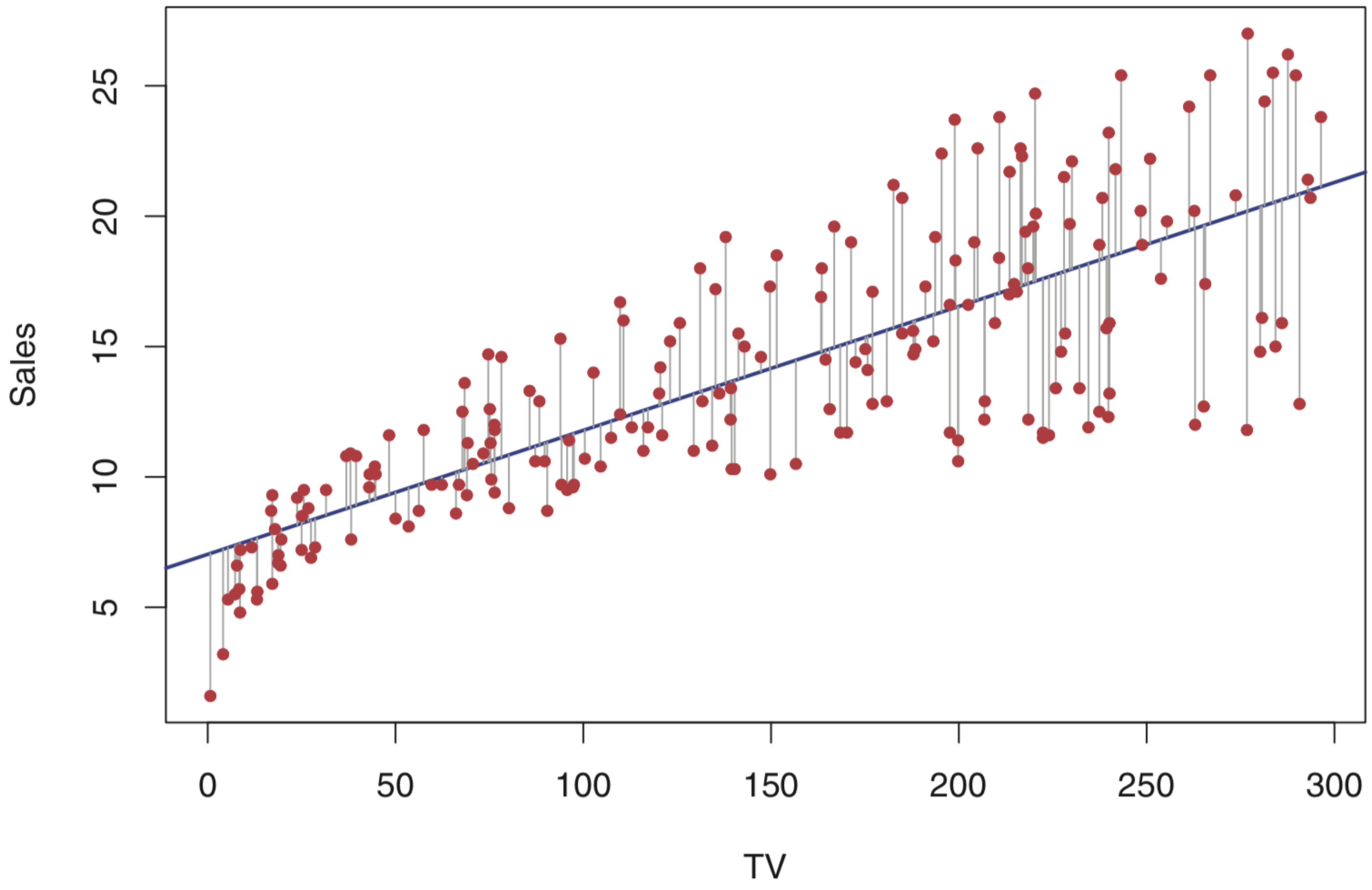
1) Which of the features/explanatory variables/predictors (x) are associated with the response variable (y)?

2) What is the relationship between x and y?

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

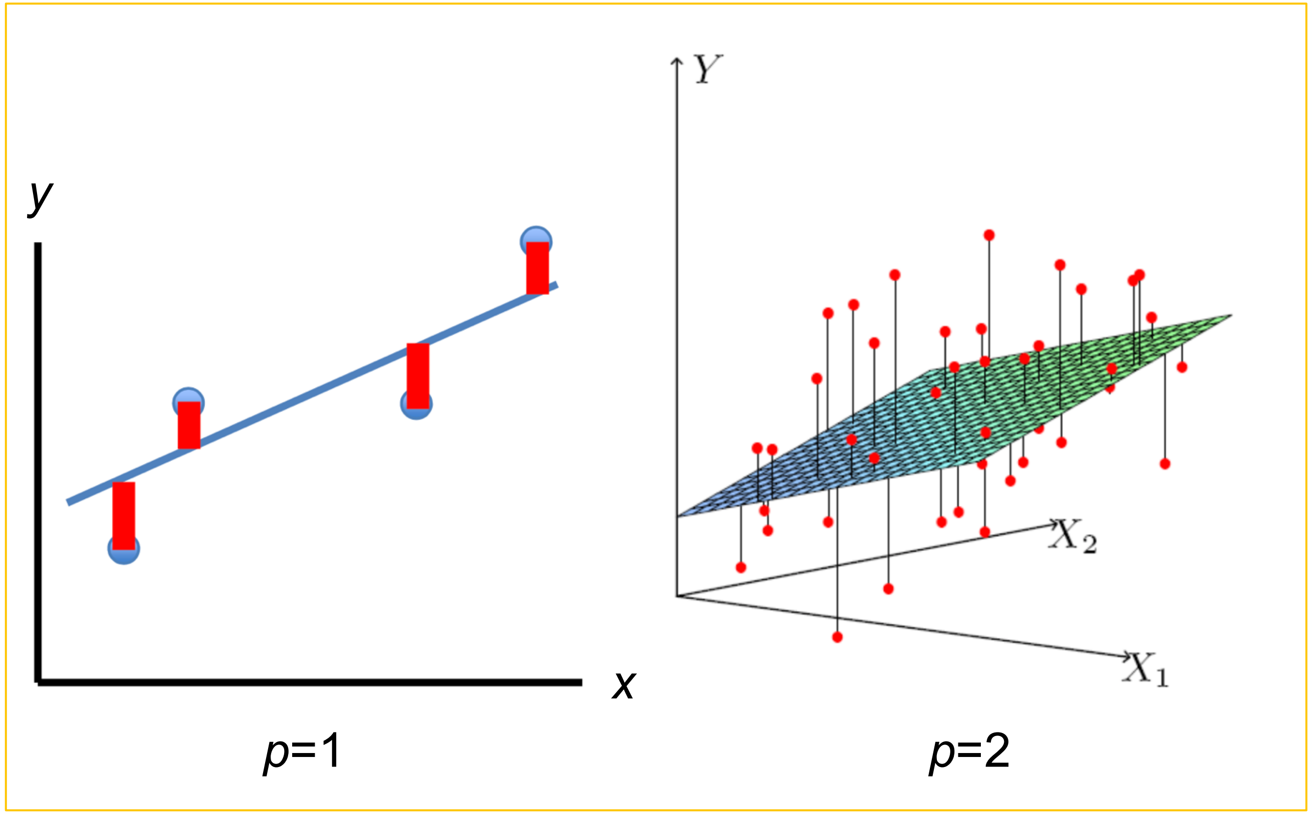
4) Is a linear model enough?

Example: predict sales from TV advertising budget



ISL: Figure 3.1

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

Why are models useful?

Linear models

Fitting a linear model (one feature)

Model complexity and evaluation

Calculus Handout

Simple linear regression

Cost function

Goal

Partial derivatives

Simple linear regression derivation

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

$$w_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

Numerator: how x and y vary **together**

Denominator: how x**varies**

$$\boxed{w_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}}$$

Variance – how much does X vary

$X = [2, 3, 4, 5, 6]$

How far is each value from the mean?

$\bar{x} = 4$

$[-2, -1, 0, 1, 2]$

How far is in total are the values from the mean?

$\sum_i (y_i - \bar{y}) = 0$

So we need to square the differences:

$$\boxed{\mathrm{Var}(X) = \frac{1}{n} \sum_i (x_i - \bar{x})^2}$$

Covariance – do X & Y move together?

For every data point, ask if x and y on the **same side** of their respective means?

$$(x_i > \bar{x}) \text{ and } (y_i > \bar{y})$$

x_i	y_i	Product
Above mean	Above mean	
Below mean	Below mean	
Above mean	Below mean	
Below mean	Above mean	

Covariance – do X & Y move together?

For every data point, ask if x and y on the **same side** of their respective means?

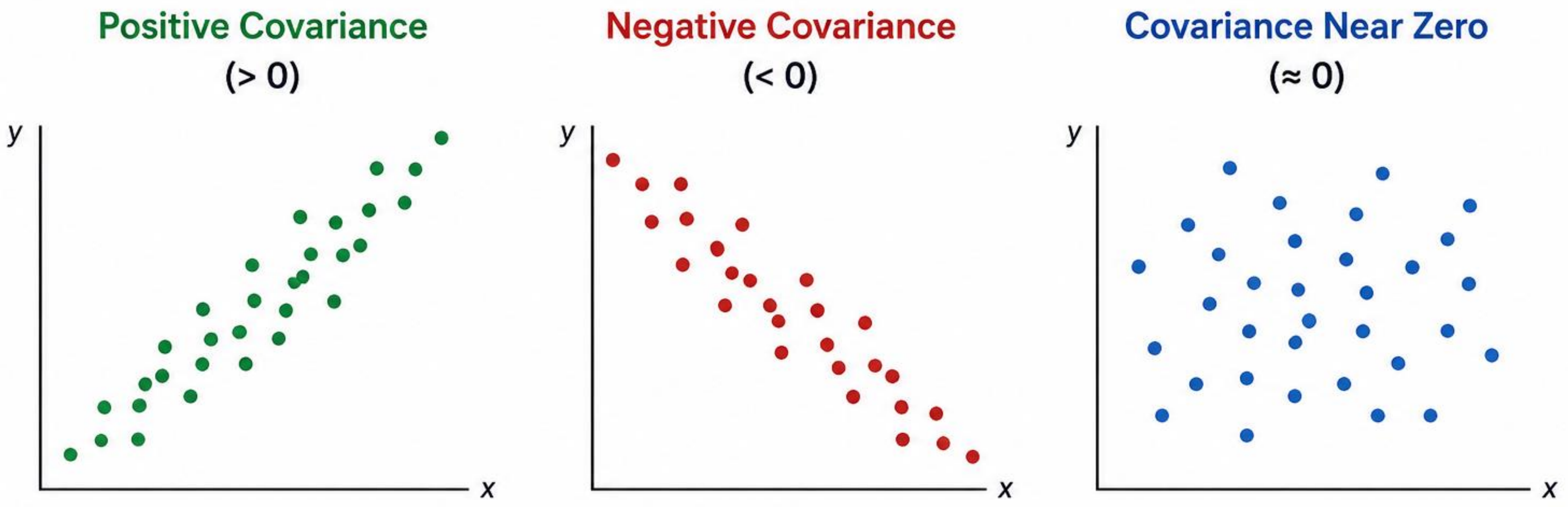
$$(x_i > \bar{x}) \text{ and } (y_i > \bar{y})$$

x_i	y_i	Product
Above mean	Above mean	$+$
Below mean	Below mean	$+$
Above mean	Below mean	$-$
Below mean	Above mean	$-$

$$\boxed{\mathrm{Cov}(X,Y) = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})}$$

What Does the Sign of Covariance Tell Us?

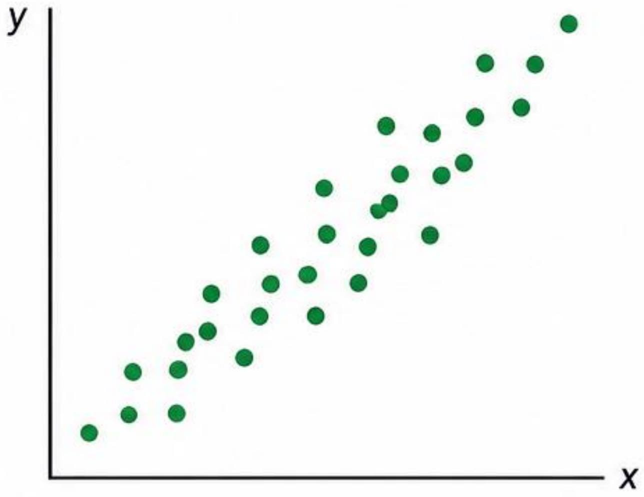
Look at the scatterplots. What sign would you expect the covariance to have?



What Does the Sign of Covariance Tell Us?

Look at the scatterplots. What sign would you expect the covariance to have?

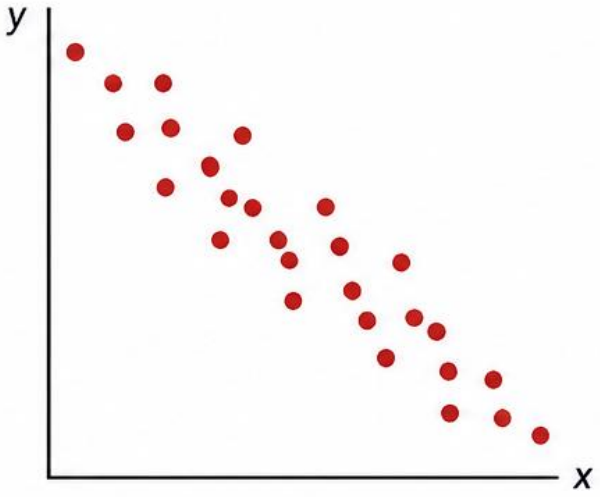
Positive Covariance
(> 0)



As X increases,
Y tends to increase.

When X is above its mean,
Y is often above its mean.
(Same side → positive products)

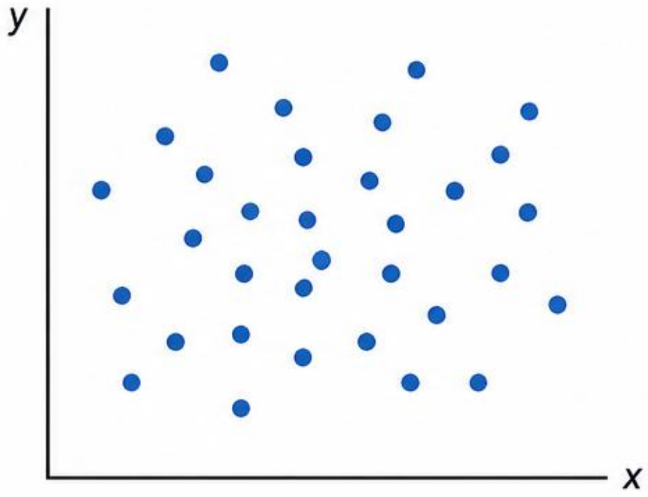
Negative Covariance
(< 0)



As X increases,
Y tends to decrease.

When X is above its mean,
Y is often below its mean.
(Opposite sides → negative products)

Covariance Near Zero
(≈ 0)



No clear linear relationship.
X and Y don't move together consistently.

Positive and negative products tend to cancel out.

Slope of best-fit line

$$\boxed{w_1 = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}}$$

$$\boxed{\text{slope} = \frac{\text{how much } X \text{ and } Y \text{ move together}}{\text{how much } X \text{ moves}}}$$

Handout 3

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?

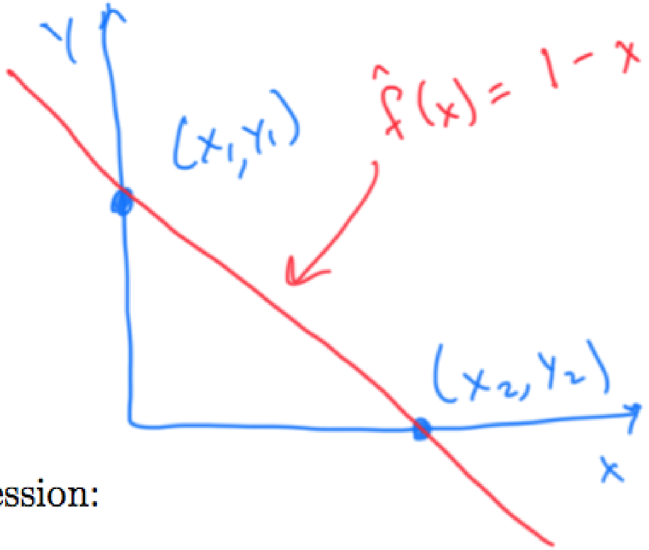
3) Can we (accurately) predict y given a new x?

4) Is a linear model enough?

Handout 3

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: $\bar{x} = \frac{1}{2}$, $\bar{y} = \frac{1}{2}$

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

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

Why are models useful?

Linear models

Fitting a linear model (one feature)

Model complexity and evaluation

Maybe a linear model is not enough

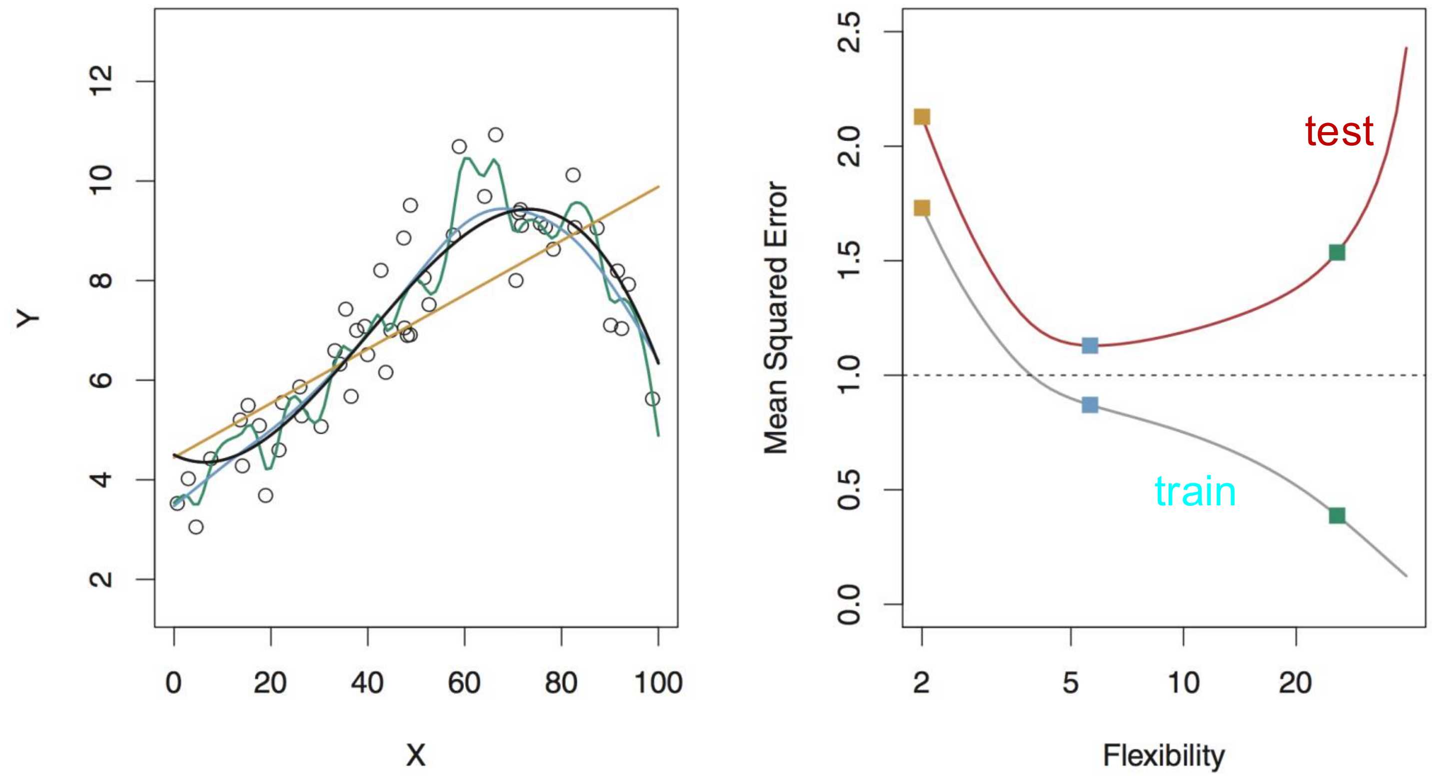
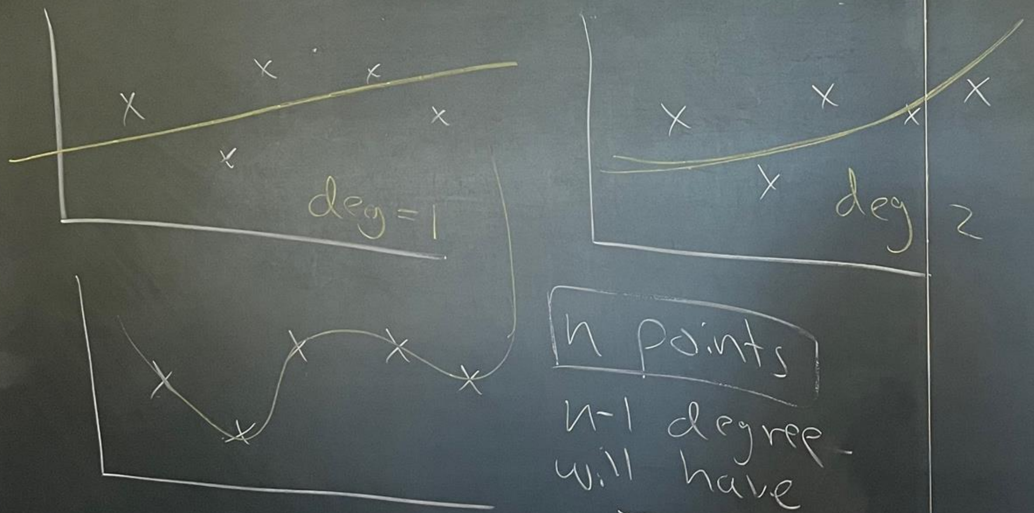


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed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



$\boxed{\text{n points}}$

n-1 degree will have

$J = 0$

Elbow Plot