Lecture 3: Linear Regression as Function Fitting

STAT 471

January 28, 2021

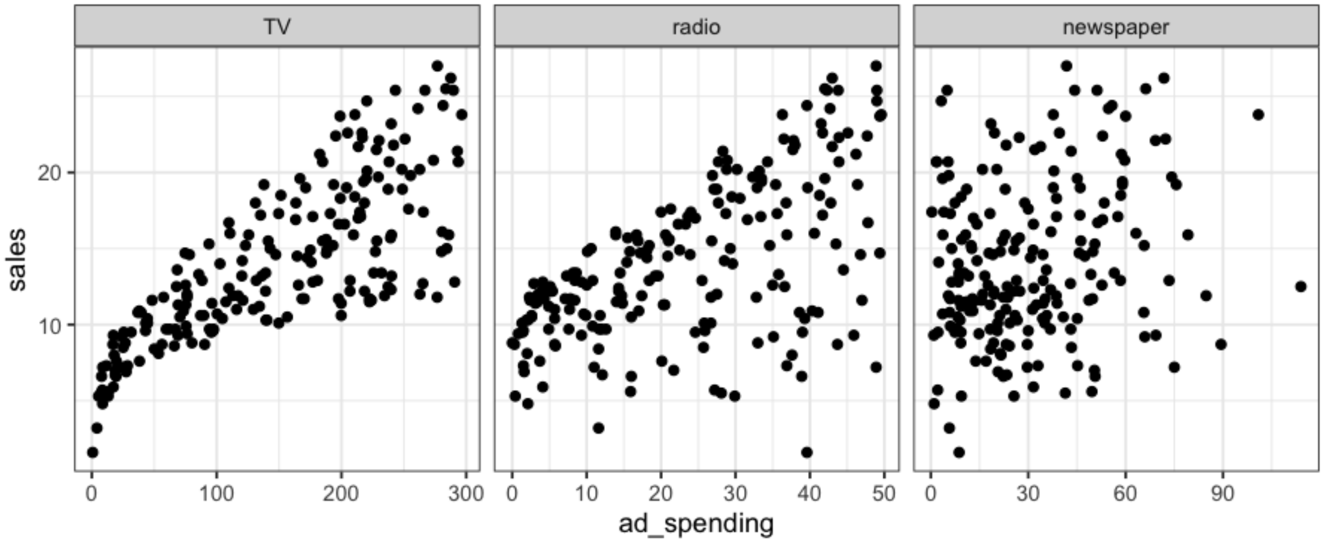
Case study: Advertising data

features (X_1, X_2, X_3) response (Y)

```
> advertising_data
# A tibble: 200 x 4
      TV radio newspaper sales
   <dbl> <dbl>     <dbl> <dbl>
 1 230.   37.8      69.2  22.1
 2  44.5  39.3      45.1  10.4
 3  17.2  45.9      69.3   9.3
 4 151.   41.3      58.5  18.5
 5 181.   10.8      58.4  12.9
 6   8.7  48.9      75     7.2
 7  57.5  32.8      23.5  11.8
 8 120.   19.6      11.6  13.2
 9   8.6   2.1       1     4.8
10 200.    2.6      21.2  10.6
# … with 190 more rows
```

How do TV, radio, and newspaper ad spending impact sales?

Data from 200 ad markets available.



The linear model

intercept

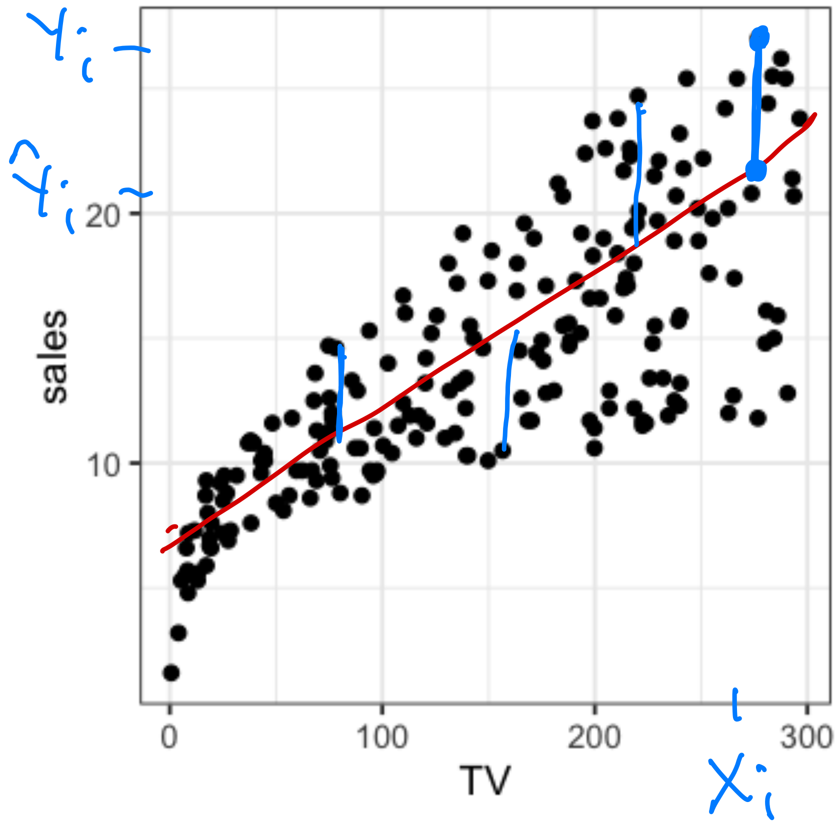
$$\text{sales} \approx \beta_0 + \beta_1 \times \text{TV} + \beta_2 \times \text{radio} + \beta_3 \times \text{newspaper}$$

Today, we will learn how to:

- Fit a linear model from training data — finding $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$
- Assess the quality of the fit
- Interpret the coefficients

Simple linear regression

$$\overset{Y}{\text{sales}} \approx \beta_0 + \beta_1 \times \overset{X}{\text{TV}}$$



$\} \; Y_i - \hat{Y}_i$ residual

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

$$Y_i \approx \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i = \hat{Y}_i$$

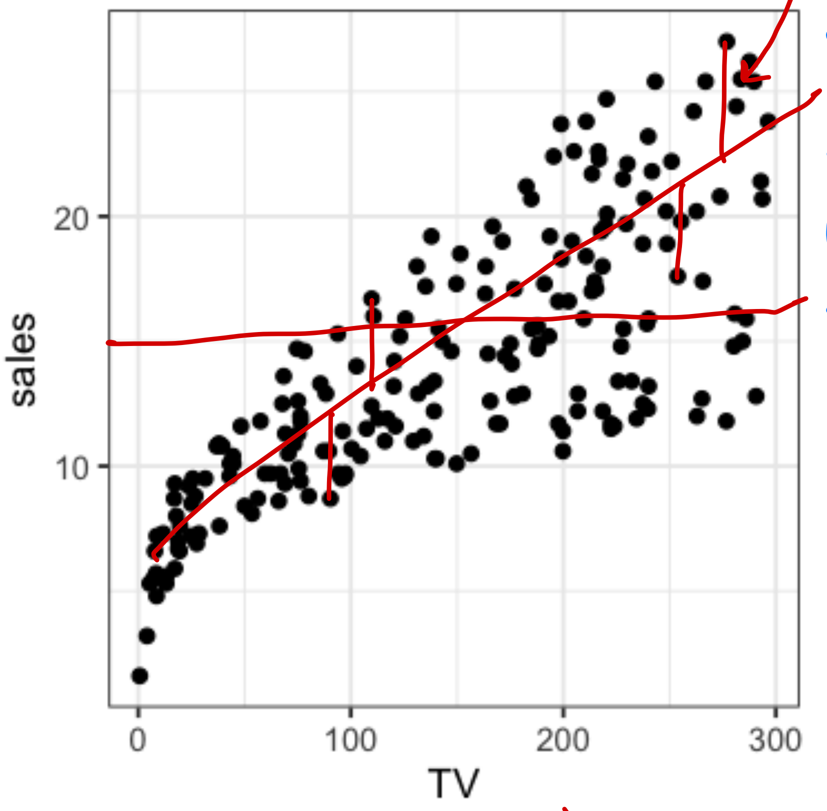
$$RSS = \sum_i (Y_i - \hat{Y}_i)^2$$

Find those $\hat{\beta}_0, \hat{\beta}_1$ that minimize the RSS.

Quality of fit

$R^2 = 0.6$

$Y_i - \hat{Y}_i$



$Y_i - \bar{Y}$

$$RSS = \sum_i (Y_i - \hat{Y}_i)^2$$

$$TSS = \sum_i (Y_i - \bar{Y})^2$$

variation explained by model

$$R^2 = \frac{TSS - RSS}{TSS}$$

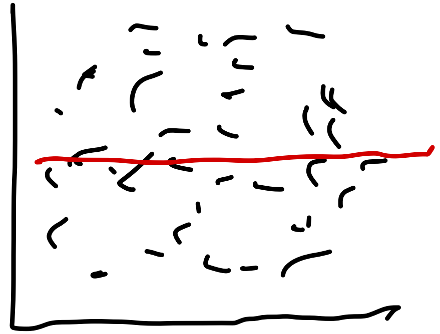
total variation in Y

Between 0 and 1.

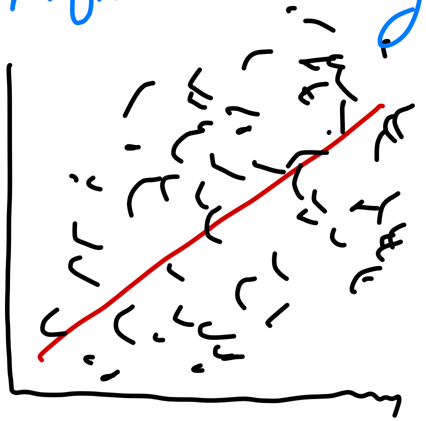
Quality of fit (examples)

Low R^2

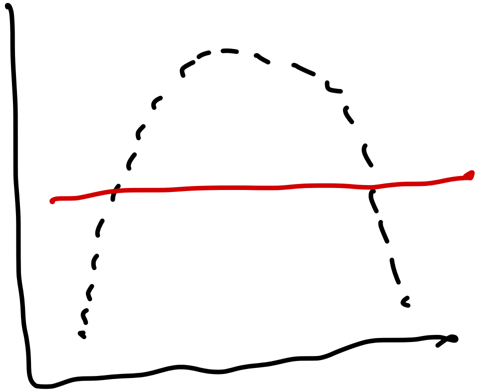
No trend



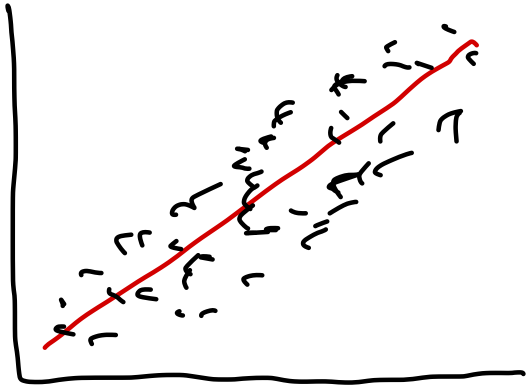
High variability



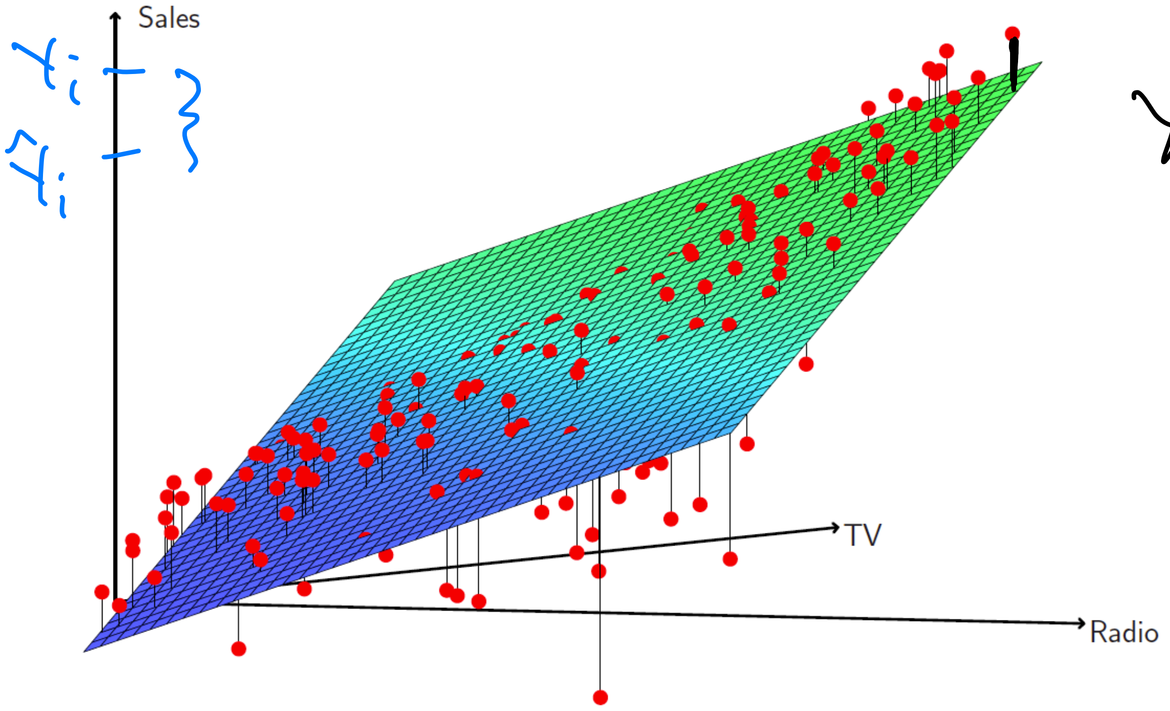
Nonlinear trend



High R^2



Multiple linear regression



$$Y_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_{i1} + \cdots + \hat{\beta}_p X_{ip} = \hat{Y}_i$$

Residuals $\quad Y_i - \hat{Y}_i$

$$RSS = \sum_i (Y_i - \hat{Y}_i)^2$$

$$R^2 = \frac{TSS - RSS}{TSS}$$

Coefficient interpretation

$$\text{sales} \approx \beta_0 + \beta_1 \times \text{TV} + \beta_2 \times \text{radio} + \beta_3 \times \text{newspaper}$$

	Coefficient
Intercept	2.939
TV	0.046
radio	0.189
newspaper	−0.001

correlation

$\hat{\beta}_3$ is the effect of newspaper when controlling for TV & radio.

Association is not causation.

$$\text{sales} \approx \beta_0 + \beta_1 \times \text{newspaper}$$

	Coefficient
Intercept	12.351
newspaper	0.055

As you increase newspaper by \$1000, you sell 55 more units.

Categorical features

Suppose we have another feature, sponsoring an event.

This feature is binary; the company either sponsors an event or not.

$X =$ "sponsored" or "not sponsored"

$$X = \begin{cases} 1 & \text{if "sponsored"} \\ 0 & \text{if "not sponsored"} \end{cases}$$

$Y = \beta_0 + \beta_1 \cdot X$

If sponsored, expect $\beta_0 + \beta_1$

If not, expect β_0.

Categorical features

Suppose the company instead wants to choose *which* event to sponsor: a football game, basketball game, or baseball game.

$$X = \begin{cases} 0 & \text{if "football"} \\ 1 & \text{if "basketball"} \\ 2 & \text{if "baseball"} \end{cases} \qquad \begin{array}{l} \rightarrow \beta_0 \\ \rightarrow \beta_0+\beta_1 \\ \rightarrow \beta_0+2\cdot\beta_1 \end{array}$$

$$\beta_0 + \beta_1 \cdot X$$

$$X_1 = \begin{cases} 1 & \text{if "football"} \\ 0 & \text{if not} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if "basketball"} \\ 0 & \text{if not} \end{cases}$$

$$\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2$$

football: $\beta_0 + \beta_1$

basketball: $\beta_0 + \beta_2$

baseball: β_0