CSCE 633: Machine Learning

Lecture 4: Linear Regression

Texas A&M University
Bobak Mortazavi

Goals For This Lecture

- Motivate a simple supervised learning problem
- Introduce a linear machine learning method (Linear regression)
- Develop a Loss Function
- Ordinary Least Squares - Optimally solve the learning problem
- Interpret model
- Understanding Accuracy and Error
- Acknowledgements: example and figure sources: James, Witten, Hastie, Tibshirani (ISLR)

Notation and Modeling

- $D = \{(x_i, y_i)\}_{i=1}^{n}$
- x_i a column vector of length p, with n samples
- y_i a scalar
- for p = 1, linear regression is fitting line to data in 2-dimensional space
- in general, linear regression is about fitting a hyperplane to a scatter of points in a p + 1 dimensional space

Notation and Modeling

- Consider the p dimensional case
- The objective is determining intercept β_0 and p slope weights $\beta_i' s$ so that for all n datapoints:

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \approx y_i$$

- Putting it into the vector form:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{bmatrix}, \dot{x}_i = \begin{bmatrix} 1 \\ x_{i1} \\ \dots \\ x_{ip} \end{bmatrix}$$

- $\dot{x_i}$ obtained by stacking a 1 on top of x_i
- Our linear equation would be

$$\beta^T x_i \approx \; y_i, i = 1, \dots, n$$

An Important Example: Advertising

- How do I make a useful Market Plan for the coming fiscal year to increase sales?
- My budget includes advertising in:
  - TV
  - Radio
  - Newspapers
- How much should I add or subtract from each to increase sales?

Important Questions to Ask

- Is there a relationship between budget and sales?
- If there is a relationship, how strong is it?
- Which of the three media contribute to sales?
- How accurately can we estimate the effect of each medium on sales?
- Is the relationship linear?
- Is there synergy among the advertising media?

Simple Linear Regression

- We want to predict y based upon a single predictor x, we want to regress y on to x:

$$\beta_0 + \beta_1 x \approx y$$

Simple Linear Regression

- We want to predict y based upon a single predictor x, we want to regress y on to x:

$$\beta_0 + \beta_1 x \approx y$$
$$\beta_0 + \beta_1 TV \approx Sales$$

Parameters

- We want to learn (trained by existing data) the parameters of the model, also known as the coefficients, β

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Where $\hat{y}$ indicates a prediction of y on the basis of x

Estimating the Coefficients

- We do not know β_0 or β_1
- So, assume we have a training set $D = \{(x_1, y_1), \cdots, (x_n, y_n)\}$
- Assume $\mathrm{n} = 200$ markets of sales and tv budget
- Goal: set $\hat{\beta}_0$ and $\hat{\beta}_1$ so we are as close to y_i from x_i for all i

Residual Sum of Squares

- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for y based on the i'th value of x
- Then the residual error is

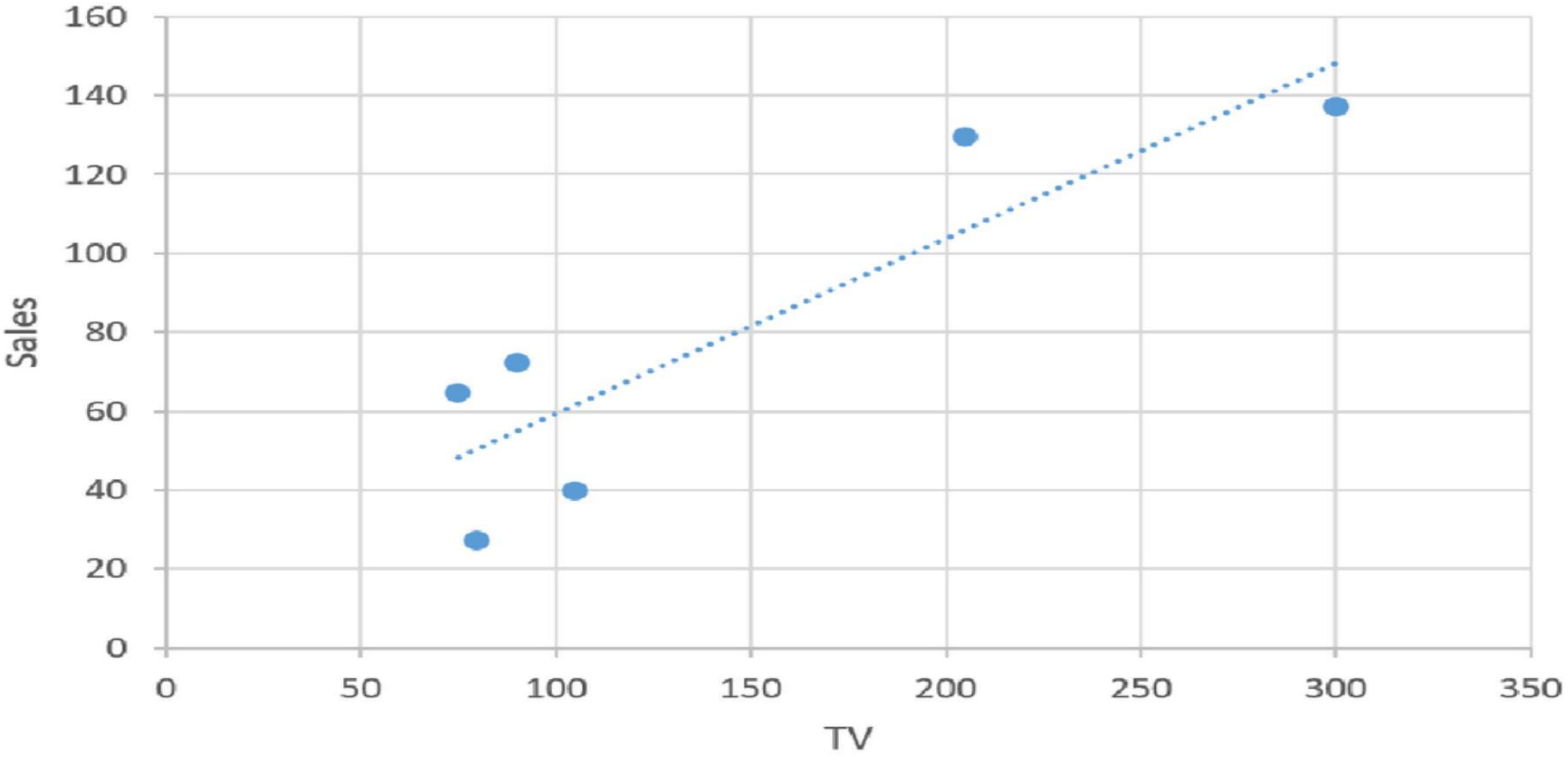
$$e_i = y_i - \hat{y}_i$$

- So, we define Residual Sum of Squares as:

$$RSS = e_1^2 + e_2^2 + \cdots + e_n^2$$

- and in least squares, the objective is minimize RSS

Residual Sum of Squares



Least Squares

The residual sum of squares

$$RSS = e_1^2 + e_2^2 + \cdots + e_n^2$$
$$= \left(y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1\right)^2 + \cdots + \left(y_N - \hat{\beta}_0 - \hat{\beta}_1 x_n\right)^2$$

Least Squares: Learning Coefficients

The residual sum of squares

$$RSS = e_1^2 + e_2^2 + \cdots + e_n^2$$
$$= \left(y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1\right)^2 + \cdots + \left(y_N - \hat{\beta}_0 - \hat{\beta}_1 x_n\right)^2$$

if RSS is our total sum of squared error, what do we need to learn?

Differentiation

To minimize RSS, need to differentiate with respect to both unknowns

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

- Calculate $\frac{\partial RSS}{\partial \hat{\beta}_0}$
- Calculate $\frac{\partial RSS}{\partial \hat{\beta}_1}$

Differentiation: $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

Differentiation: $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-1)$$

Differentiation: $\widehat{\beta}_0$

$$RSS = \sum_{\mathrm{i}=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-1)$$

$$= -2\sum_{\mathrm{i}=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)$$

Differentiation: $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-1)$$

$$= -2\sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)$$

$$= -2\sum_{\mathrm{i}=1}^{n} y_i + 2\sum_{\mathrm{i}=1}^{n} \hat{\beta}_0 + 2\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i$$

Differentiation: $\widehat{\beta}_0$

$$RSS = \sum_{\text{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = \sum_{\text{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-1)$$

$$= -2\sum_{\text{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)$$

$$= -2\sum_{\text{i}=1}^{n} y_i + 2\sum_{\text{i}=1}^{n} \hat{\beta}_0 + 2\hat{\beta}_1 \sum_{\text{i}=1}^{n} x_i$$

Note: $\bar{y} = \frac{1}{n}\sum_{\text{i}=1}^{n} y_i$ is the sample mean

Differentiation: $\widehat{\beta}_0$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-1)$$

$$= -2\sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)$$

$$= -2\sum_{\mathrm{i}=1}^{n} y_i + 2\sum_{\mathrm{i}=1}^{n} \hat{\beta}_0 + 2\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i$$

Note: $\bar{y} = \frac{1}{n}\sum_{\mathrm{i}=1}^{n} y_i$ is the sample mean

$$= -2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x}$$

Differentiation: $\widehat{\boldsymbol{\beta}}_0$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-1)$$

$$= -2\sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)$$

$$= -2\sum_{\mathrm{i}=1}^{n} y_i + 2\sum_{\mathrm{i}=1}^{n} \hat{\beta}_0 + 2\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i$$

Note: $\bar{y} = \frac{1}{n}\sum_{\mathrm{i}=1}^{n} y_i$ is the sample mean

$$= -2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x}$$

To minimize, set $\frac{\partial RSS}{\partial \hat{w}_0} = 0$

Differentiation: $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x}$$

To minimize, set $\frac{\partial RSS}{\partial \hat{w}_0} = 0$

$$-2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x} = 0$$

Differentiation: $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x}$$

To minimize, set $\frac{\partial RSS}{\partial \hat{w}_0} = 0$

$$-2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x} = 0$$

$$2n\hat{\beta}_0 = 2n\bar{y} - 2n\hat{\beta}_1\bar{x}$$

Differentiation: $\hat{\beta}_0$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x}$$

To minimize, set $\frac{\partial RSS}{\partial \hat{w}_0} = 0$

$$-2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x} = 0$$

$$2n\hat{\beta}_0 = 2n\bar{y} - 2n\hat{\beta}_1\bar{x}$$

$$\cancel{2n}\hat{\beta}_0 = \cancel{2n}\bar{y} - \cancel{2n}\hat{\beta}_1\bar{x}$$

Differentiation: $\widehat{\beta}_0$

$$RSS = \sum_{\text{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x}$$

To minimize, set $\frac{\partial RSS}{\partial \hat{\beta}_0} = 0$

$$-2n\bar{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1\bar{x} = 0$$

$$2n\hat{\beta}_0 = 2n\bar{y} - 2n\hat{\beta}_1\bar{x}$$

$$\cancel{2n}\hat{\beta}_0 = \cancel{2n}\bar{y} - \cancel{2n}\hat{\beta}_1\bar{x}$$

$$\hat{\beta}_0^* = \bar{y} - \hat{\beta}_1\bar{x}$$

Differentiation: $\widehat{\beta}_1$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-x_i)$$

Differentiation: $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$

$$RSS = \sum_{\mathrm{i}=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-x_i)$$

$$= -2\sum_{\mathrm{i}=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(x_i)$$

Differentiation: $\widehat{\beta}_1$

$$RSS = \sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = \sum_{\mathrm{i}=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-x_i)$$

$$= -2\sum_{\mathrm{i}=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(x_i)$$

set equal to 0:

$$-2\sum_{\mathrm{i}=1}^{n} y_i x_i + 2\hat{\beta}_0 \sum_{\mathrm{i}=1}^{n} x_i + 2\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i^2 = 0$$

Differentiation: $\widehat{\beta}_1$

$$-2\sum_{\mathrm{i}=1}^{n} y_i x_i + 2\hat{\beta}_0 \sum_{\mathrm{i}=1}^{n} x_i + 2\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i^2 = 0$$

$$= -\cancel{2}\sum_{\mathrm{i}=1}^{n} y_i x_i + \cancel{2}\hat{\beta}_0 \sum_{\mathrm{i}=1}^{n} x_i + \cancel{2}\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i^2 = 0$$

Differentiation: $\widehat{\beta}_1$

$$-2\sum_{\mathrm{i}=1}^{n} y_i x_i + 2\hat{\beta}_0 \sum_{\mathrm{i}=1}^{n} x_i + 2\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i^2 = 0$$

$$= -\cancel{2}\sum_{\mathrm{i}=1}^{n} y_i x_i + \cancel{2}\hat{\beta}_0 \sum_{\mathrm{i}=1}^{n} x_i + \cancel{2}\hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i^2 = 0$$

$$= -\sum_{\mathrm{i}=1}^{n} y_i x_i + (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{\mathrm{i}=1}^{n} x_i + \hat{\beta}_1 \sum_{\mathrm{i}=1}^{n} x_i^2 = 0$$

Differentiation: $\widehat{\beta}_1$

$$-2\sum_{i=1}^{n} y_i x_i + 2\hat{\beta}_0 \sum_{i=1}^{n} x_i + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$= -\cancel{2}\sum_{i=1}^{n} y_i x_i + \cancel{2}\hat{\beta}_0 \sum_{i=1}^{n} x_i + \cancel{2}\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$= -\sum_{i=1}^{n} y_i x_i + (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$-\sum_{i=1}^{n} y_i x_i + \bar{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} {x_i}^2$$

Differentiation: $\widehat{\beta}_1$

$$-2\sum_{i=1}^{n} y_i x_i + 2\hat{\beta}_0\sum_{i=1}^{n} x_i + 2\hat{\beta}_1\sum_{i=1}^{n} x_i^2 = 0$$

$$= -\cancel{2}\sum_{i=1}^{n} y_i x_i + \cancel{2}\hat{\beta}_0\sum_{i=1}^{n} x_i + \cancel{2}\hat{\beta}_1\sum_{i=1}^{n} x_i^2 = 0$$

$$= -\sum_{i=1}^{n} y_i x_i + (\bar{y} - \hat{\beta}_1\bar{x})\sum_{i=1}^{n} x_i + \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = 0$$

$$-\sum_{i=1}^{n} y_i x_i + \bar{y}\sum_{i=1}^{n} x_i - \hat{\beta}_1\bar{x}\sum_{i=1}^{n} x_i + \hat{\beta}_1\sum_{i=1}^{n} {x_i}^2$$

$$\hat{\beta}_1^* = \frac{\bar{y}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i x_i}{\bar{x}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}$$

Differentiation: $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$

$$\hat{\beta}_1^* = \frac{\bar{y}\ \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i x_i}{\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1^* = \frac{\bar{y}\ \bar{x} n - \sum_{i=1}^{n} y_i x_i}{\bar{x}^2 \mathrm{n} - \sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1^* = \frac{\sum_{i=1}^{n} y_i x_i - \bar{y}\ \bar{x} n}{\sum_{i=1}^{n} x_i^2 - \bar{x}^2 n}$$

Differentiation: $\widehat{\beta}_1$

$$\sum_{i=1}^{n} y_i x_i - \bar{y}\,\bar{x}\,n$$

$$\sum_{i=1}^{n} y_i x_i - \bar{y}\,\bar{x}\,n - \bar{y}\,\bar{x}\,n + \bar{y}\,\bar{x}\,n$$

$$\sum_{i=1}^{n} y_i x_i - \bar{y}\,\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + \bar{y}\,\bar{x}\,n$$

$$\sum_{i=1}^{n} y_i x_i - \bar{y}\,\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + \bar{y}\,\bar{x}\sum_{i=1}^{n} 1$$

$$\sum_{i=1}^{n} y_i x_i - \bar{y}\,\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \bar{y}\,\bar{x}$$

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \bar{y} x_i - \sum_{i=1}^{n} \bar{x} y_i + \sum_{i=1}^{n} \bar{y}\,\bar{x}$$

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

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

Differentiation: $\widehat{\beta}_1$

$$\hat{\beta}_1^* = \frac{\bar{y}\ \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i x_i}{\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1^* = \frac{\bar{y}\ \bar{x} n - \sum_{i=1}^{n} y_i x_i}{\bar{x}^2 \mathrm{n} - \sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1^* = \frac{\sum_{i=1}^{n} y_i x_i - \bar{y}\ \bar{x} n}{\sum_{i=1}^{n} x_i^2 - \bar{x}^2 n}$$

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

Differentiation: $\widehat{\beta}_1$

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

Differentiation: $\widehat{\boldsymbol{\beta}}_1$

$$\hat{\beta}_1^* = \frac{\bar{y}\ \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i x_i}{\bar{x}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1^* = \frac{\bar{y}\ \bar{x}n - \sum_{i=1}^{n} y_i x_i}{\bar{x}^2\mathrm{n} - \sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1^* = \frac{\sum_{i=1}^{n} y_i x_i - \bar{y}\ \bar{x}n}{\sum_{i=1}^{n} x_i^2 - \bar{x}^2 n}$$

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

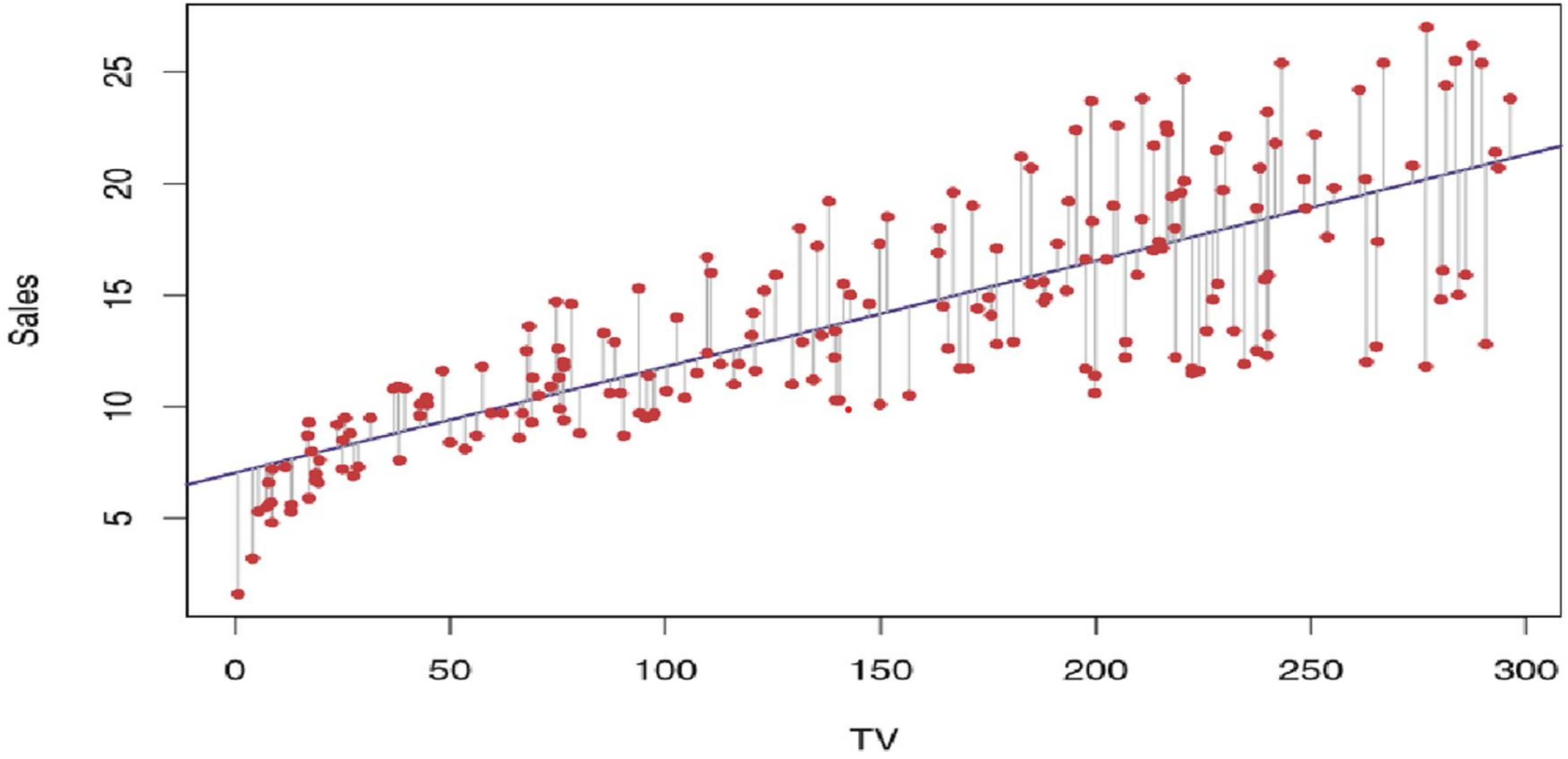
$$\hat{\beta}_1^* = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Optimal Coefficients: $\widehat{\boldsymbol{\beta}}_0, \widehat{\boldsymbol{\beta}}_1$

$$\hat{\beta}_0^* = \bar{y} - \hat{\beta}_1^* \bar{x}$$

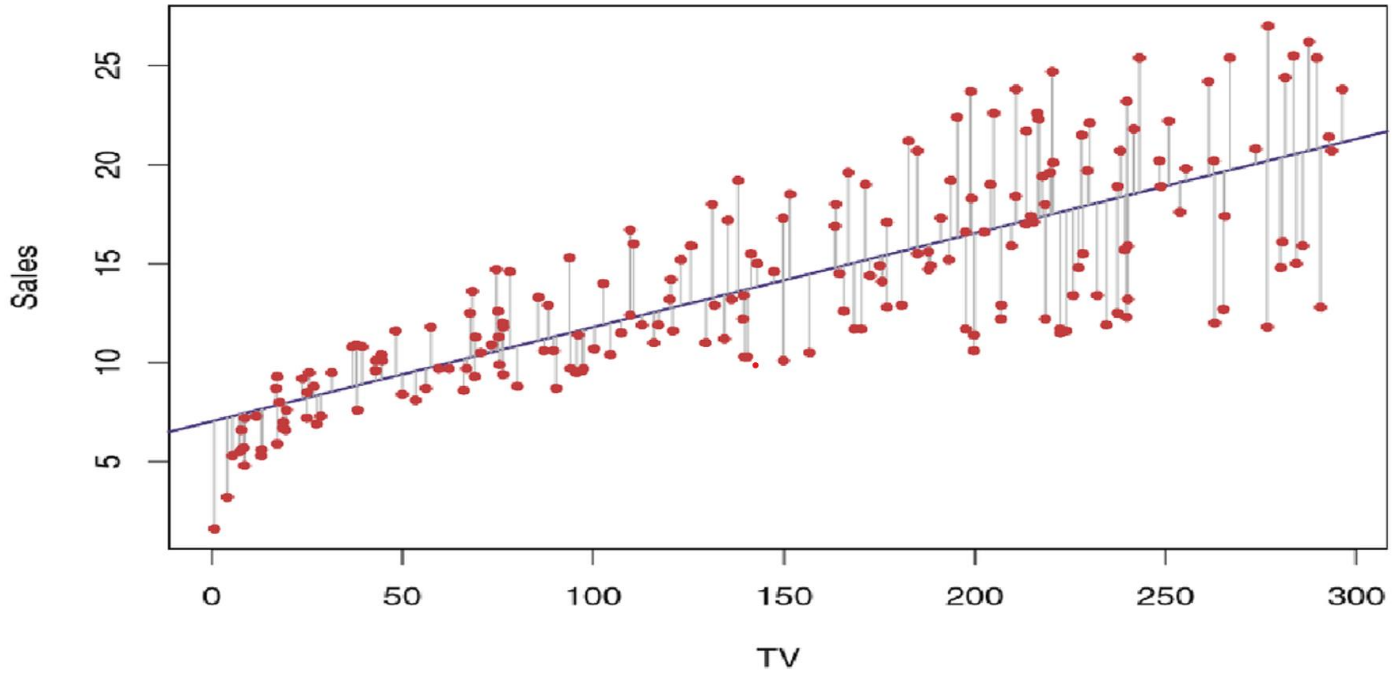
$$\hat{\beta}_1^* = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Advertising Solution



- $\hat{\beta}_0 = 7.03$
- $\hat{\beta}_1 = 0.0475$
- Source: ISLR

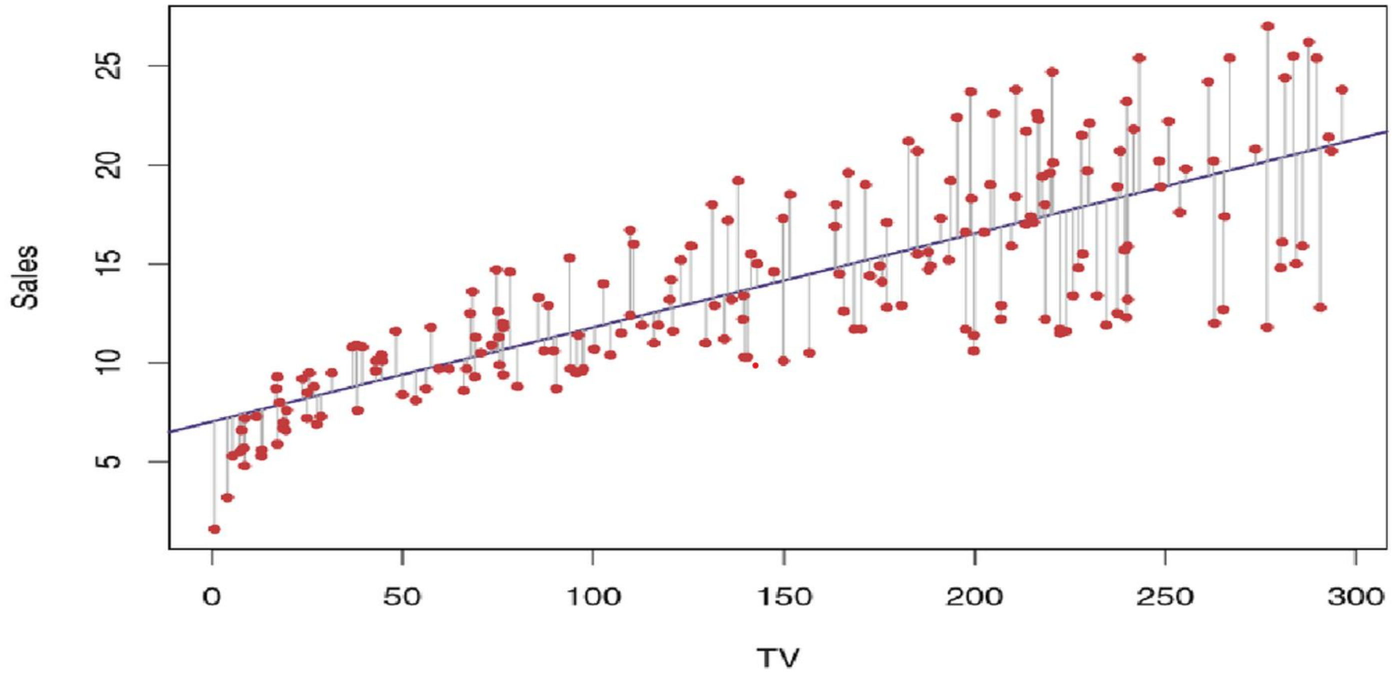
Advertising Solution



$\hat{\beta}_0 = 7.03$ and $\hat{\beta}_1 = 0.0475$. If we had no TV advertising, how many units would we sell? What if we had \$1000 budgeted for TV?

A. 703, 475 + 703
B. 7.03, 47.5 + 7.03
C. 47.5 + 7.03, 7.03
D. 475 + 703, 703

Advertising Solution



$\hat{\beta}_0 = 7.03$ and $\hat{\beta}_1 = 0.0475$. If we had no TV advertising, how many units would we sell? What if we had \$1000 budgeted for TV?

A. 703, 475 + 703

B. 7.03, 47.5 + 7.03

C. 47.5 + 7.03, 7.03

D. 475 + 703, 703

Takeaways

- Understanding key notation
- Important questions to ask for supervised learning problem
- Ordinary Least Squares
- Simple Linear Regression
- Optimizing RSS
- Next Time: Interpret model and Understanding Accuracy and Error