

Lecture 5

Simple Linear Regression

DSC 40A, Fall 2024

Agenda

- Simple linear regression.
- Minimizing mean squared error for the simple linear model.

Question 🤔

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!

If the direct link doesn't work, click the "🤔 Lecture Questions"
link in the top right corner of dsc40a.com.

Finding the best linear model

- **Goal:** Out of all **linear** functions $\mathbb{R} \rightarrow \mathbb{R}$, find the function H^* with the smallest mean squared error.
 - Linear functions are of the form $H(x) = w_0 + w_1x$.
 - They are defined by a slope (w_1) and intercept (w_0).

- That is, H^* should be the linear function that minimizes

$$R_{\text{sq}}(H) = \frac{1}{n} \sum_{i=1}^n (y_i - H(x_i))^2$$

- We chose squared loss, since it's the easiest to minimize.

Minimizing mean squared error for the simple linear model

- Our goal is to find the linear hypothesis function $H^*(x)$ that minimizes empirical risk:

$$R_{\text{sq}}(H) = \frac{1}{n} \sum_{i=1}^n (y_i - H(x_i))^2$$

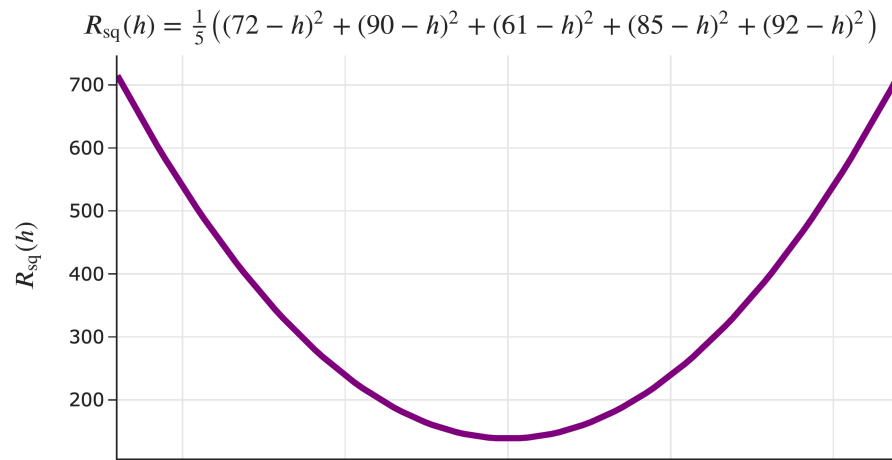
- Plugging in the linear hypothesis $H(x) = w_0 + w_1x$, we can re-write R_{sq} as a function of w_0 and w_1 :

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^n (y_i - (w_0 + w_1x_i))^2$$

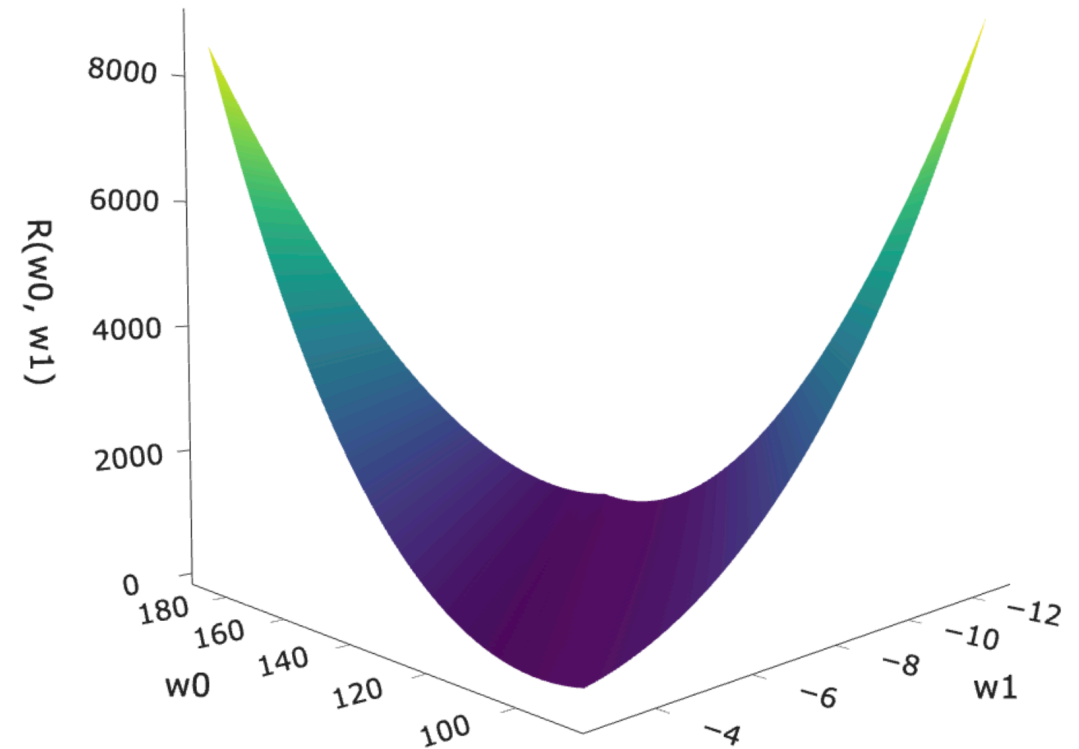
- How do we find the parameters w_0^* and w_1^* that minimize $R_{\text{sq}}(w_0, w_1)$?

Loss surface

For the constant model, the graph of $R_{\text{sq}}(h)$ looked like a parabola.



What does the graph of $R_{\text{sq}}(w_0, w_1)$ look like for the simple linear regression model?



Minimizing mean squared error for the simple linear model

Minimizing multivariate functions

- Our goal is to find the parameters w_0^* and w_1^* that minimize mean squared error:

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

- R_{sq} is a function of two variables: w_0 and w_1 .
- To minimize a function of multiple variables:
 - Take partial derivatives with respect to each variable.
 - Set all partial derivatives to 0.
 - Solve the resulting system of equations.
 - Ensure that you've found a minimum, rather than a maximum or saddle point (using the [second derivative test](#) for multivariate functions).

Example

Find the point (x, y, z) at which the following function is minimized.

$$f(x, y) = x^2 - 8x + y^2 + 6y - 7$$

Minimizing mean squared error

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

To find the w_0^* and w_1^* that minimize $R_{\text{sq}}(w_0, w_1)$, we'll:

1. Find $\frac{\partial R_{\text{sq}}}{\partial w_0}$ and set it equal to 0.
2. Find $\frac{\partial R_{\text{sq}}}{\partial w_1}$ and set it equal to 0.
3. Solve the resulting system of equations.

Question 🤔

Answer at q.dsc40a.com

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

Which of the following is equal to $\frac{\partial R_{\text{sq}}}{\partial w_0}$?

- A. $\frac{1}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i))$
- B. $-\frac{1}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i))$
- C. $-\frac{2}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i)) x_i$
- D. $-\frac{2}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i))$

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

$$\frac{\partial R_{\text{sq}}}{\partial w_0} =$$

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

$$\frac{\partial R_{\text{sq}}}{\partial w_1} =$$

Strategy

We have a system of two equations and two unknowns (w_0 and w_1):

$$-\frac{2}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i)) = 0 \quad -\frac{2}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i)) x_i = 0$$

To proceed, we'll:

1. Solve for w_0 in the first equation.

The result becomes w_0^* , because it's the "best intercept."

2. Plug w_0^* into the second equation and solve for w_1 .

The result becomes w_1^* , because it's the "best slope."

Solving for w_0^*

$$-\frac{2}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i)) = 0$$

Solving for w_1^*

$$-\frac{2}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i)) x_i = 0$$

Least squares solutions

We've found that the values w_0^* and w_1^* that minimize R_{sq} are:

$$w_1^* = \frac{\sum_{i=1}^n (y_i - \bar{y})x_i}{\sum_{i=1}^n (x_i - \bar{x})x_i} \qquad w_0^* = \bar{y} - w_1^* \bar{x}$$

where:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \qquad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

These formulas work, but let's re-write w_1^* to be a little more symmetric.

An equivalent formula for w_1^*

Claim:

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

Proof:

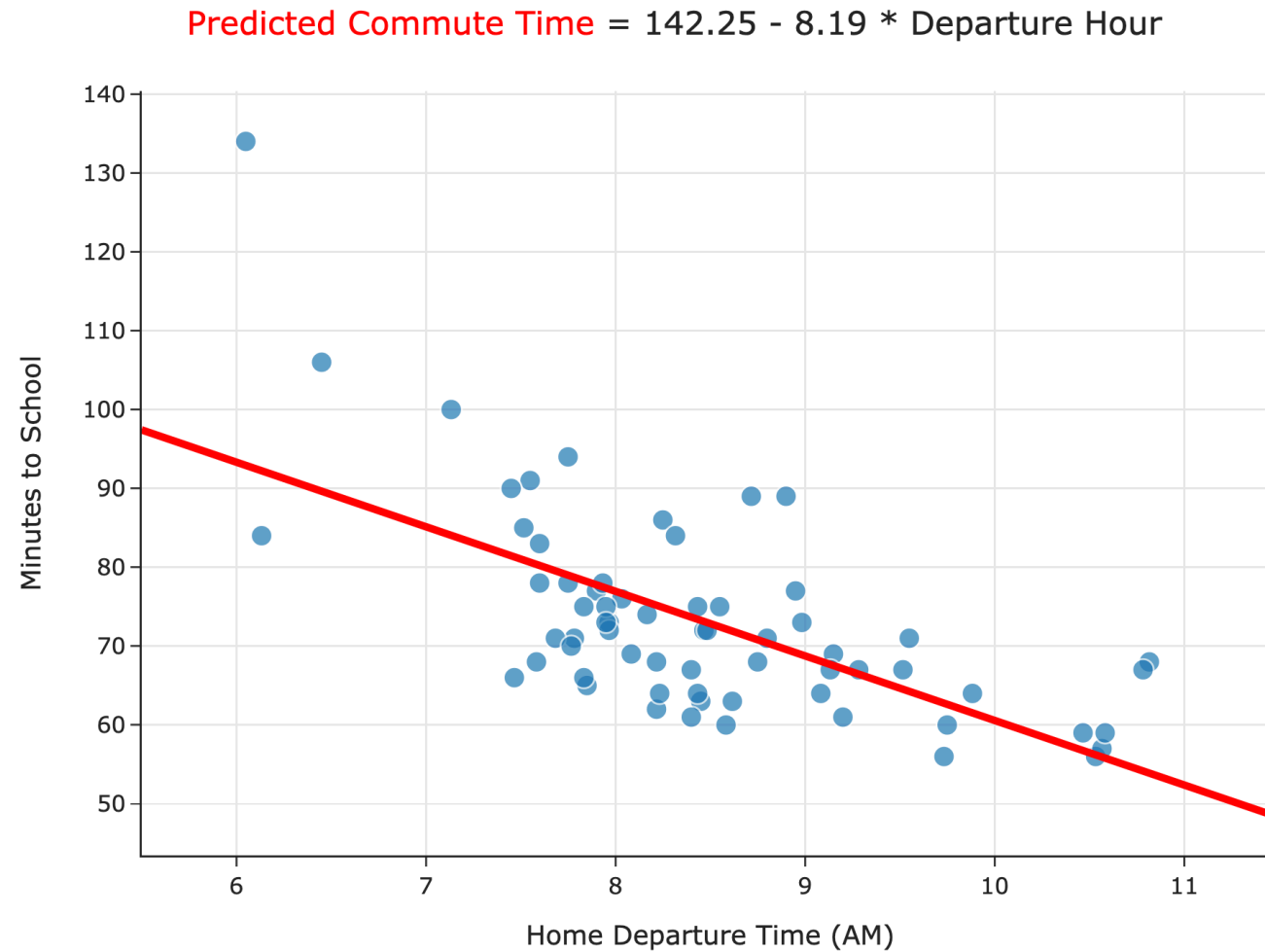
Least squares solutions

- The **least squares solutions** for the intercept w_0 and slope w_1 are:

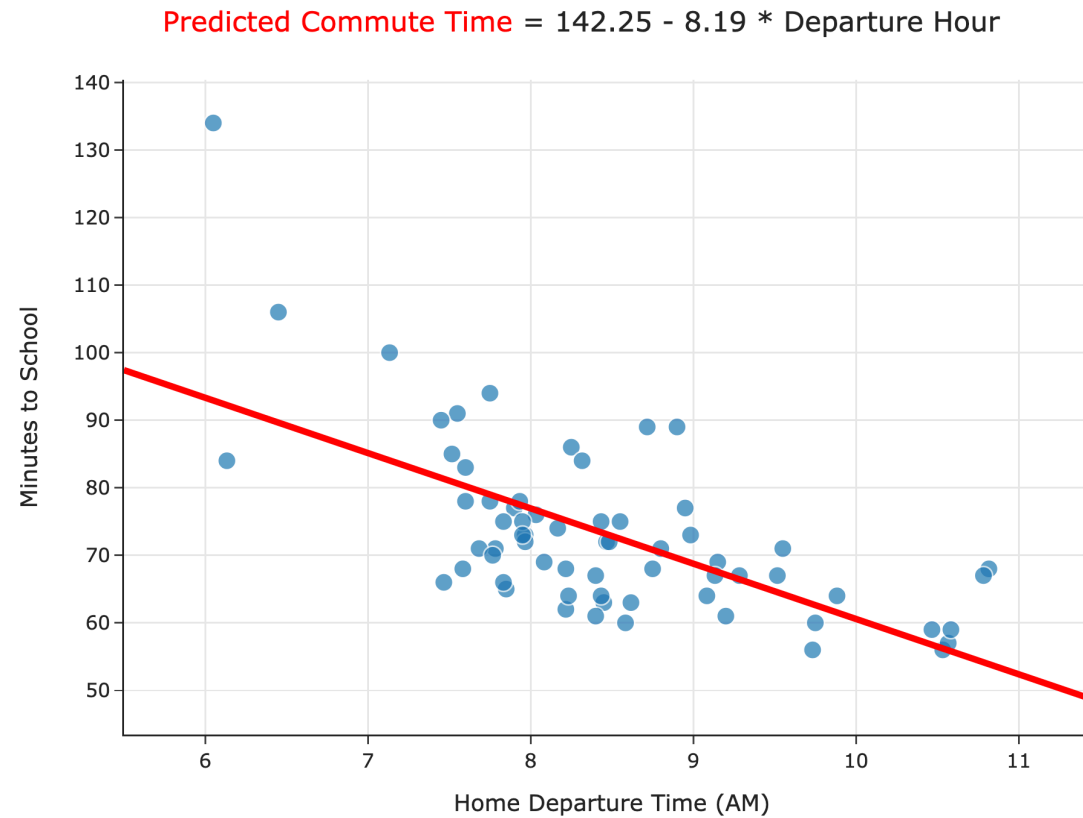
$$w_1^* = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \qquad w_0^* = \bar{y} - w_1^* \bar{x}$$

- We say w_0^* and w_1^* are **optimal parameters**, and the resulting line is called the **regression line**.
- The process of minimizing empirical risk to find optimal parameters is also called "**fitting to the data**."
- To make predictions about the future, we use $H^*(x) = w_0^* + w_1^*x$.

Let's test these formulas out in code! Follow along [here](#).



Causality



Can we conclude that leaving later **causes** you to get to school quicker?

What's next?

We now know how to find the optimal slope and intercept for linear hypothesis functions. Next, we'll:

- See how the formulas we just derived connect to the formulas for the slope and intercept of the regression line we saw in DSC 10.
 - They're the same, but we need to do a bit of work to prove that.
- Learn how to interpret the slope of the regression line.
- Discuss *causality*.
- Learn how to build regression models with **multiple inputs**.
 - To do this, we'll need linear algebra!