Lecture 7 – Simple Linear Regression (continued)

DSC 40A, Fall 2022 @ UC San Diego

Dr. Truong Son Hy, with help from **many others**

Announcements

- **Don't forget to submit your homework!**
- Look at the readings linked on the course website!
- Groupwork Relsease Day: Thursday afternoon
  Groupwork Submission Day: Monday midnight
  Homework Release Day: Friday after lecture
  Homework Submission Day: Friday before lecture
- See `dsc40a.com/calendar` for the Office Hours schedule.

Agenda

- Recap of prediction rules.
- Simple linear regression.

How do we predict someone's salary?

After collecting salary data, we...

1. Choose a loss function.

2. Find the best prediction by minimizing empirical risk.

- So far, we've been predicting future salaries without using any information about the individual (e.g. GPA, years of experience, number of LinkedIn connections).

- **New focus:** How do we incorporate this information into our prediction-making process?

Features

A **feature** is an attribute – a piece of information.

- **Numerical**: age, height, years of experience
- **Categorical**: college, city, education level
- **Boolean**: knows Python?, had internship?

Think of features as columns in a DataFrame (i.e. table).

	YearsExperience	Age	FormalEducation	Salary
0	6.37	28.39	Master's degree (MA, MS, M.Eng., MBA, etc.)	120000.0
1	0.35	25.78	Some college/university study without earning ...	120000.0
2	4.05	31.04	Bachelor's degree (BA, BS, B.Eng., etc.)	70000.0
3	18.48	38.78	Bachelor's degree (BA, BS, B.Eng., etc.)	185000.0
4	4.95	33.45	Master's degree (MA, MS, M.Eng., MBA, etc.)	125000.0

Variables

- The features, x, that we base our predictions on are called **predictor variables**.
- The quantity, y, that we're trying to predict based on these features is called the **response variable**.
- We'll start by predicting salary based on years of experience.

Prediction rules

- We believe that salary is a function of experience.
- In other words, we think that there is a function H such that:
$$\text{salary} \approx H(\text{years of experience})$$
- H is called a **hypothesis function** or **prediction rule**.
- **Our goal**: find a good prediction rule, H.

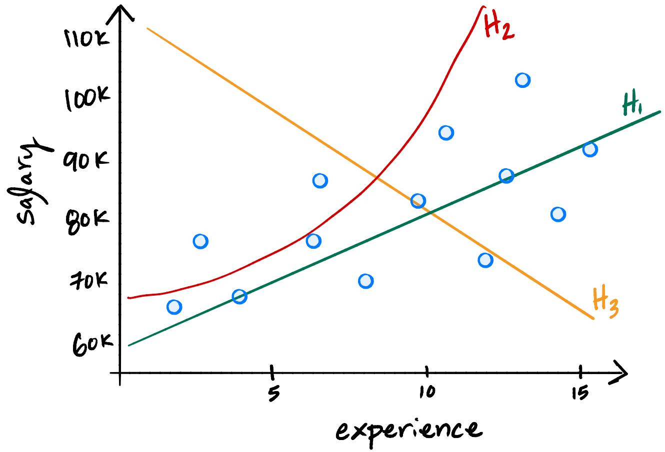
Comparing predictions

- How do we know which prediction rule is best: H_1, H_2, H_3?
- We gather data from n people. Let x_i be experience, y_i be salary:

$$\begin{matrix} (\text{Experience}_1, \text{Salary}_1) \\ (\text{Experience}_2, \text{Salary}_2) \\ \dots \\ (\text{Experience}_n, \text{Salary}_n) \end{matrix} \quad \rightarrow \quad \begin{matrix} (x_1, y_1) \\ (x_2, y_2) \\ \dots \\ (x_n, y_n) \end{matrix}$$

- See which rule works better on data.

Example



Quantifying the quality of a prediction rule H

- Our prediction for person i's salary is $H(x_i)$.
- As before, we'll use a **loss function** to quantify the quality of our predictions.
  - Absolute loss: $\left|y_i - H(x_i)\right|$.
  - Squared loss: $\left(y_i - H(x_i)\right)^2$.
- We'll use squared loss, since it's differentiable.
- Using squared loss, the **empirical risk** (mean squared error) of the prediction rule H is:

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

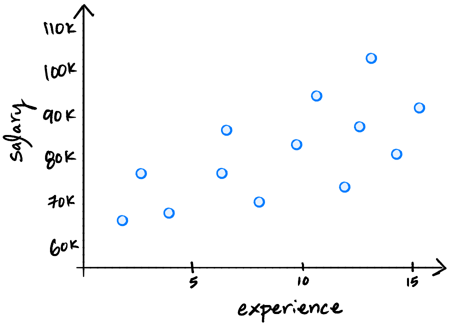
Finding the best prediction rule

- **Goal:** out of all functions $\mathbb{R} \to \mathbb{R}$, find the function H^* with the smallest mean squared error.

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

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

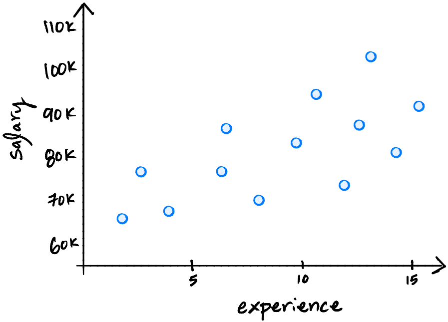
- There's a problem.



Discussion Question

Given the data above, is there a prediction rule H which has **zero** mean squared error?

a) Yes   b) No



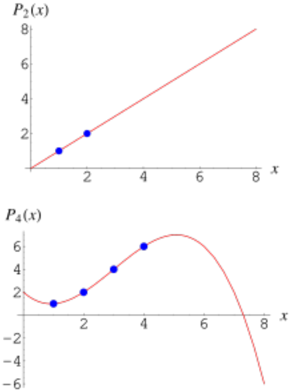
Discussion Question

Given the data above, is there a prediction rule H which has **zero** mean squared error?

a) Yes b) No

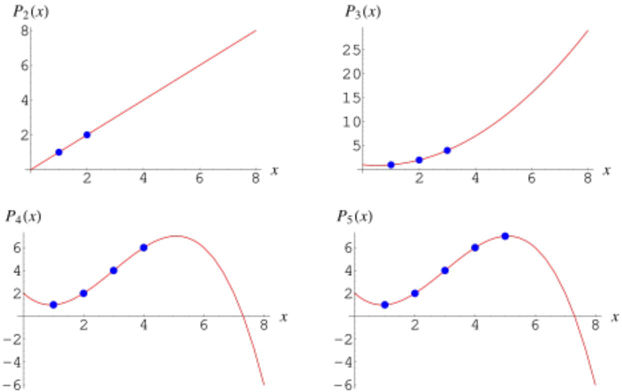
Answer: Yes

Lagrange interpolation (polynomial)



The degree of the polynomial is exactly the number of data points

Lagrange interpolation (polynomial)



The degree of the polynomial is exactly the number of data points

Problem

- We can make mean squared error very small, even zero!
- But the function will be weird.
- This is called **overfitting**.
- Remember our real goal: make good predictions on data **we haven't seen**.

Solution

- Don't allow H to be just any function.
- Require that it has a certain form.
- Examples:
  - Linear: $H(x) = w_0 + w_1 x$.
  - Quadratic: $H(x) = w_0 + w_1 x_1 + w_2 x^2$.
  - Exponential: $H(x) = w_0 e^{w_1 x}$.
  - Constant: $H(x) = w_0$.

Finding the best linear prediction rule

- **Goal:** out of all linear functions $\mathbb{R} \to \mathbb{R}$, find the function H^* with the smallest mean squared error.
  - Linear functions are of the form $H(x) = w_0 + w_1 x$.
  - They are defined by a slope (w_1) and intercept (w_0).
- That is, H^* should be the linear function that minimizes

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

- This problem is called **least squares regression**.
  - "Simple linear regression" refers to linear regression with a single predictor variable.

Minimizing mean squared error for the linear prediction rule

Minimizing the mean squared error

- The MSE is a function R_{sq} of a function H.

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

- But since H is linear, we know $H(x_i) = w_0 + w_1 x_i$.

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

- Now R_{sq} is a function of w_0 and w_1.
- We call w_0 and w_1 **parameters**.
  - Parameters define our prediction rule.

Updated goal

- Find the slope w_1^* and intercept w_0^* that minimize the MSE, $R_{sq}(w_0, w_1)$:

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

- Strategy: multivariable calculus.

Recall: the gradient

- If $f(x, y)$ is a function of two variables, the **gradient** of f at the point (x_0, y_0) is a **vector** of **partial derivatives**:

$$\nabla f(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) \\ \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}$$

- **Key Fact #1**: The derivative is to the tangent line as the gradient is to the tangent plane.

- **Key Fact #2**: The gradient points in the direction of the biggest increase.

- **Key Fact #3**: The gradient is zero at critical points.

Strategy

To minimize $R(w_0, w_1)$: compute the gradient, set it equal to zero, and solve.

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

Discussion Question

Choose the expression that equals $\frac{\partial R_{\text{sq}}}{\partial w_0}$.

a) $\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_i)\right)$

b) $-\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_i)\right)$

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

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

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

Discussion Question

Choose the expression that equals $\frac{\partial R_{sq}}{\partial w_0}$.

a) $\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_i)\right)$

b) $-\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_i)\right)$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Discussion Question

Choose the expression that equals $\frac{\partial R_{sq}}{\partial w_1}$.

a) $\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_i)\right)$

b) $-\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_i)\right)$

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

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

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

Discussion Question

Choose the expression that equals $\frac{\partial R_{sq}}{\partial w_1}$.

a) $\frac{1}{n}\sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right)$

b) $-\frac{1}{n}\sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right)$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Strategy

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

1. Solve for w_0 in first equation.
    - The result becomes w_0^*, since it is the “best intercept”.

2. Plug w_0^* into second equation, solve for w_1.
    - The result becomes w_1^*, since it is the “best slope”.

Solve for w_0^*

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

Solve for w_0^*

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

$$\Leftrightarrow \sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right) = 0$$

Solve for w_0^*

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

$$\Leftrightarrow \sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right) = 0$$

$$\Leftrightarrow -n w_0 + \sum_{i=1}^{n}\left(y_i - w_1 x_i\right) = 0$$

Solve for w_0^*

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

$$\Leftrightarrow \sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right) = 0$$

$$\Leftrightarrow -nw_0 + \sum_{i=1}^{n}\left(y_i - w_1 x_i\right) = 0$$

$$\Leftrightarrow w_0 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - w_1 x_i\right)$$

Solve for w_0^*

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

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

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

$$\Leftrightarrow w_0 = \frac{1}{n}\sum_{i=1}^{n}\big(y_i - w_1 x_i\big)$$

$$\Leftrightarrow w_0 = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - w_1\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \overline{y} - w_1\overline{x}$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1} y_i$.

Solve for w_1^*

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

Solve for w_1^*

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

$$\Leftrightarrow \sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right)x_i = 0$$

Solve for w_1^*

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

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

$$\Leftrightarrow \sum_{i=1}^{n} y_i x_i - w_0 \sum_{i=1}^{n} x_i - w_1 \sum_{i=1}^{n} x_i^2 = 0$$

Replace $w_0 = \overline{y} - w_1 \overline{x}$, we have:

Solve for w_1^*

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

$$\Leftrightarrow \sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right)x_i = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} y_i x_i - w_0 \sum_{i=1}^{n} x_i - w_1 \sum_{i=1}^{n} x_i^2 = 0$$

Replace $w_0 = \overline{y} - w_1\overline{x}$, we have:

$$\Leftrightarrow \sum_{i=1}^{n} y_i x_i - (\overline{y} - w_1\overline{x}) \sum_{i=1}^{n} x_i - w_1 \sum_{i=1}^{n} x_i^2 = 0$$

Solve for w_1^*

We have:

$$\sum_{i=1}^{n} y_i x_i - \bar{y} \sum_{i=1}^{n} x_i + w_1 \bar{x} \sum_{i=1}^{n} x_i - w_1 \sum_{i=1}^{n} x_i^2 = 0$$

Solve for w_1^*

We have:

$$\sum_{i=1}^{n} y_i x_i - \overline{y} \sum_{i=1}^{n} x_i + w_1 \overline{x} \sum_{i=1}^{n} x_i - w_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} (y_i - \overline{y}) x_i - w_1 \sum_{i=1}^{n} (x_i - \overline{x}) x_i = 0$$

Solve for w_1^*

We have:

$$\sum_{i=1}^{n} y_i x_i - \overline{y} \sum_{i=1}^{n} x_i + w_1 \overline{x} \sum_{i=1}^{n} x_i - w_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} (y_i - \overline{y}) x_i - w_1 \sum_{i=1}^{n} (x_i - \overline{x}) x_i = 0$$

$$\Leftrightarrow w_1 = \frac{\sum_{i=1}^{n} (y_i - \overline{y}) x_i}{\sum_{i=1}^{n} (x_i - \overline{x}) x_i}$$

Least squares solutions

- We've found that the values w_0^* and w_1^* that minimize the function $R_{sq}(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right)^2$ are

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

  where

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

- Let's re-write the slope w_1^* to be a bit more symmetric.

Key fact

The **sum of deviations from the mean** for any dataset is 0.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \sum_{i=1}^{n}(y_i - \bar{y}) = 0$$

Proof:

Key fact

The **sum of deviations from the mean** for any dataset is 0.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \sum_{i=1}^{n}(y_i - \bar{y}) = 0$$

Proof:
From definition, we have:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Key fact

The **sum of deviations from the mean** for any dataset is 0.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \sum_{i=1}^{n}(y_i - \bar{y}) = 0$$

Proof:
From definition, we have:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\Leftrightarrow n\overline{x} = \sum_{i=1}^{n} x_i$$

Key fact

The **sum of deviations from the mean** for any dataset is 0.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \sum_{i=1}^{n}(y_i - \bar{y}) = 0$$

Proof:
From definition, we have:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\Leftrightarrow\ n\overline{x} = \sum_{i=1}^{n} x_i \ \Leftrightarrow\ \sum_{i=1}^{n}\overline{x} = \sum_{i=1}^{n} x_i$$

Key fact

The **sum of deviations from the mean** for any dataset is 0.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \sum_{i=1}^{n}(y_i - \bar{y}) = 0$$

Proof:
From definition, we have:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\Leftrightarrow n\overline{x} = \sum_{i=1}^{n} x_i \Leftrightarrow \sum_{i=1}^{n}\overline{x} = \sum_{i=1}^{n} x_i \Leftrightarrow \sum_{i=1}^{n}(x_i - \overline{x}) = 0.$$

Similarly for $\overline{y}$.

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:

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: Because $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, we have:

$$-\overline{x}\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

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: Because $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, we have:

$$-\overline{x}\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

Because $\sum_{i=1}^{n}(y_i - \overline{y}) = 0$, we have:

$$-\overline{x}\sum_{i=1}^{n}(y_i - \overline{y}) = 0$$

Equivalent formula for w_1^*

Proof (continued):
We have:

$$w_1 = \frac{\sum_{i=1}^n (y_i - \overline{y})x_i}{\sum_{i=1}^n (x_i - \overline{x})x_i} = \frac{\sum_{i=1}^n (y_i - \overline{y})x_i + 0}{\sum_{i=1}^n (x_i - \overline{x})x_i + 0}$$

Thus:

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

Therefore:

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

Least squares solutions

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

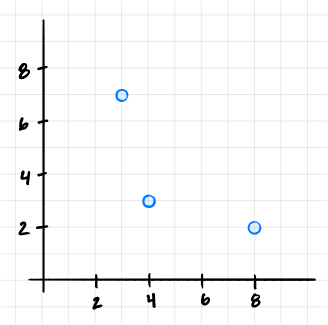
$$w_1^* = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad w_0^* = \bar{y} - w_1 \bar{x}$$

  - We also say that w_0^* and w_1^* are **optimal parameters**.

- To make predictions about the future, we use the prediction rule

$$H^*(x) = w_0^* + w_1^* x$$

Example



$\bar{x} =$

$\bar{y} =$

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

$$w_0^* = \bar{y} - w_1 \bar{x} =$$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
3	7				
4	3				
8	2				

Optional homework: Write a Python/MATLAB/C++ program to compute w_1^* and w_0^* given any data $\{(x_i, y_i)\}_{i=1}^{n}$.

Summary

Summary, next time

- We introduced the linear prediction rule, $H(x) = w_0 + w_1 x$.
- To determine the best choice of slope (w_1) and intercept (w_0), we chose the squared loss function $(y_i - H(x_i))^2$ and minimized empirical risk $R_{sq}(w_0, w_1)$:

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

- After solving for w_0^* and w_1^* through partial differentiation, we have a prediction rule $H^*(x) = w_0^* + w_1^* x$ that we can use to make predictions about the future.
- **Next time**: Revisiting correlation from DSC 10. Revisiting gradient descent. Introducing a linear algebraic formulation of linear regression.