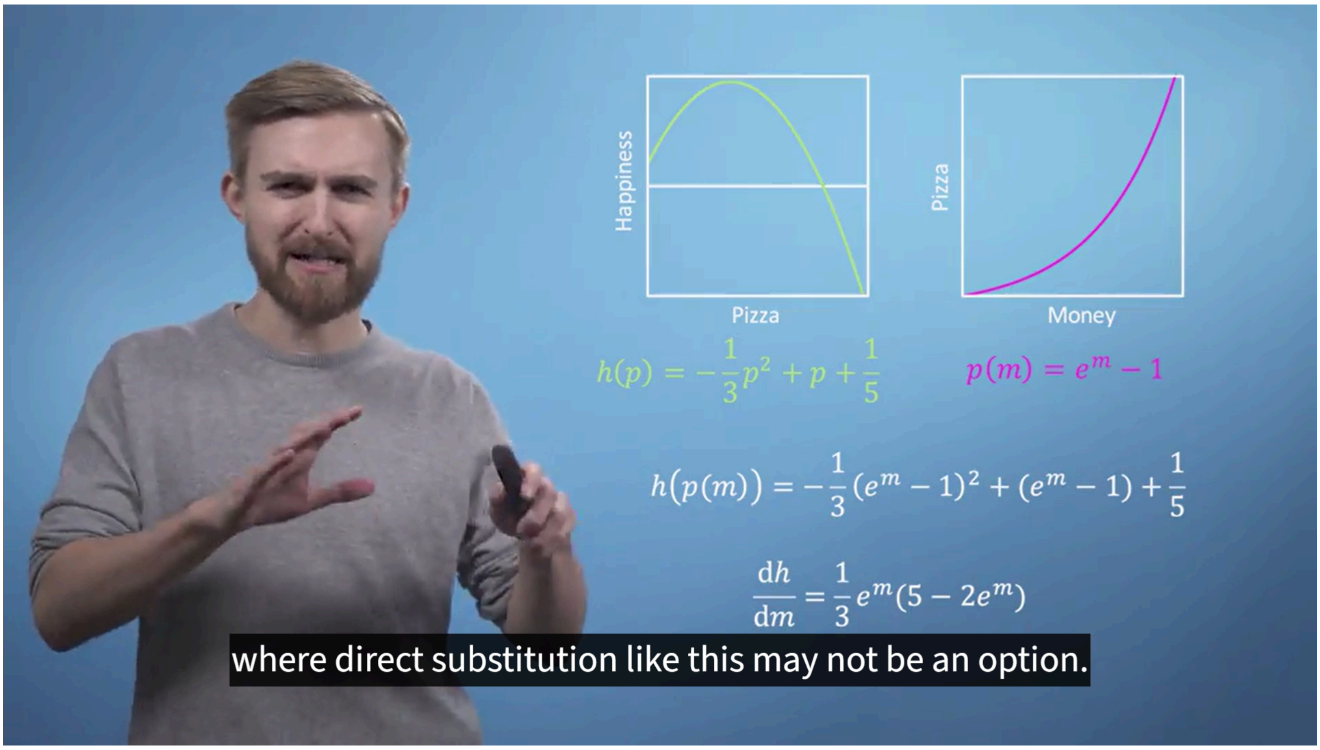
06: The Chain Rule

Chapter Goal: To build a strong intuition for the Chain Rule, the rule for differentiating composite (nested) functions.

---

1. Background: Functions Inside Functions (Composition)

- **Problem:** We often encounter functions whose input is the output of another function. This is called a **composite function** or a nested function.
- **Intuitive Example (The Pizza Chain):**
  - $h(p)$: **Happiness** (`h`) is a function of the **pizza** (`p`) eaten.
  - $p(m)$: The amount of **pizza** (`p`) you can buy is a function of the **money** (`m`) you have.
  - **Combined Function:** We have a chain of relationships: **Money** → **Pizza** → **Happiness**. We can write this as $h(p(m))$.



- **The Key Question:**

  > "How fast does my happiness (`h`) change if my money (`m`) changes a little bit?"

  - We want to find $\frac{dh}{dm}$.

---

2. Two Ways to Solve the Problem

Method #1: Direct Substitution (Brute Force)

1. Substitute the formula for $p(m)$ into every `p` in the formula for $h(p)$.
2. This will result in one very complicated function `h` that depends only on `m`.
3. Differentiate this complicated function directly.

- **Disadvantage:** This can be very difficult and algebraically intensive if the functions are complex.

Method #2: The Chain Rule (The Elegant Way)

- **Core Idea:** We can use the derivatives of each "link" in the chain separately.
- We know two things:
  1. How fast happiness changes with respect to pizza ($\frac{dh}{dp}$).
  2. How fast pizza changes with respect to money ($\frac{dp}{dm}$).
- **The "Domino Effect" Logic:**
  The total change in happiness due to a change in money is:
  `(The rate of change of h with respect to p) * (The rate of change of p with respect to m)`

---

3. "Aha!" Moment: The Intuitive Leibniz Notation

If we write the "domino effect" logic above using derivative (Leibniz) notation, we get:

> $$\frac{dh}{dm} = \frac{dh}{dp} \cdot \frac{dp}{dm}$$

- **Notational Intuition:** Although this is not a standard fraction cancellation, visually the `dp` on top and the `dp` on the bottom look like they can "cancel out," leaving `dh/dm`.
- This is a very powerful mnemonic for remembering the Chain Rule.

---

4. Applying the Chain Rule to the Pizza Example

1. **Find the Derivative of Each "Link":**
   - $h(p) = -\frac{1}{3}p^2 + p + \frac{1}{5} \implies \frac{dh}{dp} = 1 - \frac{2}{3}p$
   - $p(m) = e^m - 1 \implies \frac{dp}{dm} = e^m$
2. **Multiply the Two Derivatives (Apply the Chain Rule):**

$$\frac{dh}{dm} = \frac{dh}{dp} \cdot \frac{dp}{dm} = \left(1 - \frac{2}{3}p\right) \cdot e^m$$

3. **Return to the Language of `m`:**

- Our result still contains `p`. We must substitute it back with $p(m) = e^m - 1$.

$$\frac{dh}{dm} = \left(1 - \frac{2}{3}(e^m - 1)\right) \cdot e^m$$

(After simplification, this will be identical to the result from the direct substitution method).

---

5. Key Message

- The Chain Rule might seem like a "long way around" for simple examples.
- **Its True Power:** In the real world and in ML, we often don't know the analytical formula for a function, but we can *measure* its derivatives (e.g., from experimental data).
- In such cases, the Chain Rule becomes extremely powerful because it allows us to combine known derivatives to find a total derivative that we don't know. This is the **core idea behind Backpropagation** in Neural Networks.

$$\frac{d}{dx}\left[\left(f(x)\right)^n\right] = n\left(f(x)\right)^{n-1} \cdot f'(x)$$

$$\frac{d}{dx}\left[f\left(g(x)\right)\right] = f'\left(g(x)\right)g'(x)$$

Tags: #mml-specialization #multivariate-calculus #derivatives #chain-rule #backpropagation