AMES Class Notes – Week Five, Day 1

Andie Creel

23$^{\text{rd}}$ September, 2024

1 Review from last week

1.1 Notation:

$$\frac{dF(x)}{dx} = F'(x) \tag{1}$$

$$\tag{2}$$

1.2 Chain rule:

Consider $F(G(X))$

$$\frac{dF}{dx} = \frac{dF}{dG}\frac{dG}{dx} \tag{3}$$

1.3 We can work with derivatives like fractions.

Therefore,

$$\frac{dF(x)}{dx} = F'(x) \implies dF(x) = F'(x)dx \tag{4}$$

This means we went from a *derivative* to a *differential*. This is an important part of notation with integrals as well. The thing to know and remember is that dx just means "small changes." We can divide by small changes (like in derivatives) or multiply by small changes (like in differentials).

1.4 Derivative rule for exponential function, e^x

:

$$\frac{de^x}{dx} = e^x \tag{5}$$

We have proportional change. This is the only function in the world where the rate of change at a point is equal to the function at that point.

1.5 Example:

Consider the function
$$f(x) = e^{ax}$$

We know
$$\frac{de^x}{dx} = e^x.$$

Let's do a u-substitution to make this simpler. Let
$$U(x) = ax.$$

Our function is now written as

$$f(x) = e^U.$$

Recall the chain rule:

$$\frac{dF(x)}{dx} = \frac{dF(x)}{dU(x)} \frac{dU(x)}{dx} \tag{6}$$

We want to know what $\frac{dF(x)}{dx}$ is. We can solve for the two simpler derivatives on the right hand side to gt this.

$$\frac{dF(x)}{dU(x)} = e^U$$

$$\frac{dU(x)}{dx} = a \implies$$

$$\frac{dF(x)}{dx} = ae^U$$

Now, plug our U-substitution back in:

$$\frac{dF(x)}{dx} = ae^{ax}$$

1.6   Example

Let

$$F(x) = ln(e^x) \tag{7}$$

We're interested in the derivative.

$$\frac{dF(x)}{dx} = 1 \tag{8}$$

Why!? Chain rule, log rule, exponential rule. First, do a u-substitution and apply the chain rule

$$U(x) = e^x \implies \tag{9}$$

$$F(x) = ln(U(x)) \implies \tag{10}$$

$$\frac{dF(x)}{dx} = \frac{dF(U(x))}{dx} = \frac{dF(x)}{dU(x)} \frac{dU(x)}{dx} \tag{11}$$

Find $\frac{dF(x)}{dU(x)}$ using log rules

$$\frac{dF(x)}{dU(x)} = \frac{1}{U(x)} \tag{12}$$

Find $\frac{dU(x)}{dx}$ using exponential rules

$$\frac{dU(x)}{dx} = e^x \tag{13}$$

Now we can plug these results back into equation 11,

$$\frac{dF(x)}{dx} = \frac{1}{U(x)} e^x \tag{14}$$

Plug our u-substitution back in,

$$= \frac{1}{e^x} e^x \tag{15}$$

$$= 1 \tag{16}$$

2 Maximum and minimum

If you're out hiking, how do you know if you're at the top of a hill? You will walk on flats at the top. We often want to know the peaks and valleys of slopes. When are we at the top? When are we at the bottom?

In applied setting we often times want to know when we can maximize and objective or minimize an objective? How much time does a critter spend vigilant to maximize it's chance of survival? How much money should a business invest in climate mitigation strategies to maximize their profit?

These questions are interesting on their own. These are the questions that motivate *why* we need to learn calculus, because these questions can be answered *using* calculus. Calculus isn't important on its own, it's important because it can help us answer important questions.

2.1 How do we know if we're going to a peak or a valley?

We can use calculus to find a peak or find a valley. Consider a mountain range that can be modeled with the equation

$$Y = F(x) = ax - bx^2 \tag{17}$$
$$a > 0 \text{ and } b > 0 \tag{18}$$

where is a peak/valley of this mountain range? At what level of x is $F(x)$ maximized or minimized?

$$\frac{dF(x)}{dx} = a - 2bx = 0 \implies \tag{19}$$
$$x = \frac{a}{2b} \tag{20}$$

How do we know if this is a maximum or minimum (peak or valley)? We can use the **second derivative**. If the second derivative is negative, it's a maximum. If the second derivative is positive, it's a minimum. (Use Eli's tricks of smiley faces and frowny faces to remember this. If the second derivative is negative, the function is a frowny. If the second derivative is positive, the function is a smiley).

A function is concave is the second derivative is negative. A function is convex if the second derivative is positive.

$$\frac{d^2 F(x)}{dx^2} = -2b. \tag{21}$$

Assuming $b > 0$ implies $-2b < 0$ therefore we have a local maximum.

Physicists do a good job keeping of functions, first derivatives and second derivatives. It's (respectively) location, speed, acceleration, jerk ("if someone's being really mean to you you can say you're being a really big third derivative" - Eli).

3 How the eqn for mean was derived

How do we find an average? We want to find an $\bar{x}$ that minimizes the total distance from $\bar{x}$, squared. Why are we squaring it? We don't care if x_i is greater that $\bar{x}$ or less than $\bar{x}$. We just want to find an $\bar{x}$ where all x_i are not that far away from $\bar{x}$.

We can define the mean as the x that minimizes the differences between itself and all other x_i observations,
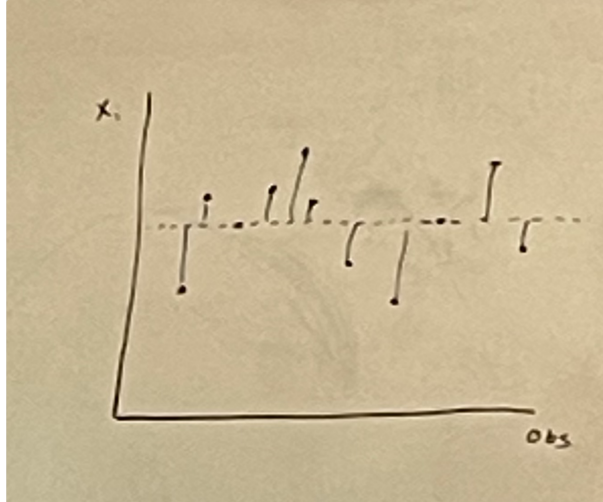
Figure 1: Dotted line is mean

$$\bar{x} = \mathbf{argmin} \sum_{i}^{N} (x_i - \bar{x})^2 \tag{22}$$

Great, we have an equation defining the mean. However, this is not very easy to compute (how do you use an argmin??). It's defined clearly, but we can't calculate it easily.

3.1   Deriving eqn for mean

Consider the function

$$F(\bar{x}) = \sum_{i}^{N} (x_i - \bar{x})^2 \tag{23}$$

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + ... + (x_N - \bar{x})^2 \tag{24}$$

Objective: Find the $\bar{x}$ that minimizes the distance from itself and all other x_i, where the distance between two points x, y is measured as $(x - y)^2$.

If we're wanting to *minimize* (getting back to the argmin) the distance from x_i to $\bar{x}$, we can take a derivative and set it equal to zero:

$$\frac{dF(\bar{x})}{d\bar{x}} = -2(x_1 - \bar{x}) - 2(x_2 - \bar{x}) - 2(x_3 - \bar{x}) - ... - 2(x_N - \bar{x}) = 0 \tag{25}$$

$$\tag{26}$$

Divide both sides by -2, then group the $\bar{x}$ into one term,

$$x_1 + x_2 + x_3 + ...x_n - N\bar{x} = 0 \tag{27}$$

$$x_1 + x_2 + x_3 + ...x_n = N\bar{x} \tag{28}$$

plug in summation notation

$$\sum_i^N x_i = N\bar{x} \implies \tag{29}$$

$$\bar{x} = \frac{1}{N} \sum_i^N x_i \tag{30}$$

This is how the **arithmetic mean** was derived. When someone wanted to find an $\bar{x}$ that was not that different from all the other x_i, they solved this problem!

To confirm that $\bar{x}$ *minimizes* the distances between all x_i and itself, you can take a second derivative and confirm that the second derivative is positive, which means the function is convex and we've found the minimum of that convex function.

3.2   Problem set hint

The hardest problem on the problem set is asking us to derive the process for linear regression.

Linear regression finds a conditional mean *aka* a mean conditioned on x, which is just a line.

In the problem set, we will want to minimize the distance between y_i and $F(x_i)$,

$$\text{distance} = (y_i - F(x_i))^2. \tag{31}$$

What's $F(x)$ going to be? The previously mentioned conditional mean, *i.e.,* a line. In the problem set, we parameterize a line that minimizes the distance between the data points and the line. Solving this problem solves for the same algorithms that computers use when they fit linear regressions.

4   Mean Value Theorem

Definition of Mean Value Theorem: For a smooth and continuous function, there is at least one point on a given interval where the derivative is equal to the average rate of change over the interval.

Intuition: The theorem says that if you draw the secant line (the straight line) connecting the point $(a, f(a))$ (where a is on the x-axis and $f(x)$ is on the y-axis) and the point $(b, f(b))$, then somewhere in the interval (a, b) there will be a point c where the tangent line (the line that touches the curve at exactly one point and has the same slope as the curve at that point) is parallel to the secant line. In other words, at some point in the interval, the function's derivative matches the average rate of change over the interval,

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{32}$$

5   Taylor Series

We know that straight lines are good approximators because a line is a conditional mean, and means are good at minimizing error.

Series are good approximators of sequences of numbers, because series are functions.

A **Taylor Series** is an extremely useful series for approximating the relationship of sequence of numbers (data is a sequence of numbers). Taylor series underpins a lot of the math we do today, particularly for any relationship that isn't linear.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ... + \frac{f^k(a)}{k!}(x-a)^k + R \qquad (33)$$

where R is the residual, or higher order terms.

5.1   0th order Taylor Series

Consider a 0th order taylor series: it would be a **mean** aka just a flat line equal to $f(a)$.

If you're only looking at the means of a dataset, that means you're doing a 0th order taylor approximation.

5.2   First order Taylor Series

Consider a first order taylor series:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a)^1 \qquad (34)$$

$$= A + B(x-a) \qquad (35)$$

$$= A - aB + Bx \qquad (36)$$

$$= m + Bx \qquad (37)$$

It's a **line** with a slope. A first order Taylor series is a linear approximation. Linear regression is a first order Taylor approximation.