AMES Class Notes – Week Five, Day 1

Andie Creel

25$^{\text{th}}$ September, 2023

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:

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

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

$$Y = F(x) = ax - bx^2 \tag{17}$$

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

$$\frac{dF(x)}{dx} = a - 2bx = 0 \implies \tag{18}$$

$$x = \frac{a}{2b} \tag{19}$$

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).

$$\frac{d^2 F(x)}{dx^2} = -2b. \tag{20}$$

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.

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 $\bar{x}$.

$$\bar{x} = \textbf{argmin} \sum_{i}^{N} (x_i - \bar{x})^2 \tag{21}$$

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.

Consider the function

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

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + ... + (x_N - \bar{x})^2 \tag{23}$$

Now, 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:

$$\frac{dF(\bar{x})}{d\bar{x}} = -2(x_1 - \bar{x}) - 2(x_2 - \bar{x}) - 2(x_3 - \bar{x}) - ... - 2(x_N - \bar{x}) = 0 \tag{24}$$

$$\tag{25}$$

Pull out $\bar{x}$,

$$x_1 + x_2 + x_3 + ...x_n - N\bar{X} = 0 \tag{26}$$
$$x_1 + x_2 + x_3 + ...x_n = N\bar{X} \tag{27}$$

plug in summation notation

$$\sum_i^N x_i = N\bar{x} \implies \tag{28}$$

$$\bar{x} = \frac{1}{N}\sum_i^N x_i \tag{29}$$

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

3.1 Problem set hint

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

Linearly regression finds a *conditional* mean. A mean conditioned on x is just a line.

Hint: This distance we'll want to be minimizing is

$$y_i - F(x_i) \tag{30}$$

what's $F(x)$? That's what we are parameterizing or estimating.

4 Partial derivatives

Consider a function that maps from $R^2 \to R^1$,

$$Z = F(x, y). \tag{31}$$

Consider a situation where we can change x with policy, but we can't change y and we want to maximize outcome Z. In this case, we would take the derivative of $F(x, y)$ w.r.t x (treat y like a constant, assume it isn't changing).

We can take a **partial derivative** which notation that are all equivalent:

$$\frac{\partial F(x, y)}{\partial x} = F_x(x, y) = F_1(x, y) \tag{32}$$

This is how we find a **relative max/min**. Conditioned on y staying constant, we can find the relative max/min (relative to that level of y).

4.1 Absolute max/min

If you want to find the absolute maximum or minimum you need to take *all* partial derivatives and set them all equal to zero and solve that system of equations. If your function mapped from $R^N \to R^1$ then you would take N partial derivatives and solve the system of N equations.