

# FAME Week Three Class Notes – Derivatives II

Andie Creel

11<sup>th</sup> November, 2025

## 1 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 frequently want to know when we can maximize an objective or minimize an objective? How much time does a critter spend vigilant to maximize its 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.

### 1.1 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

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

where is a peak/valley of this mountain range? At what level of  $x$  is  $F(x)$  maximized or minimized? We can see when the derivative of the function equals zero, then solve for  $x$ . This value of  $x$  tells us where the function is "flat".

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

How do we know if  $x = a/(2b)$  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 if the second derivative is negative. A function is convex if the second derivative is positive.

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

Assuming  $b > 0$  implies  $-2b < 0$  therefore we have a local maximum because the second derivative is negative.

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).

## 2 Deriving the equation for taking the mean/average

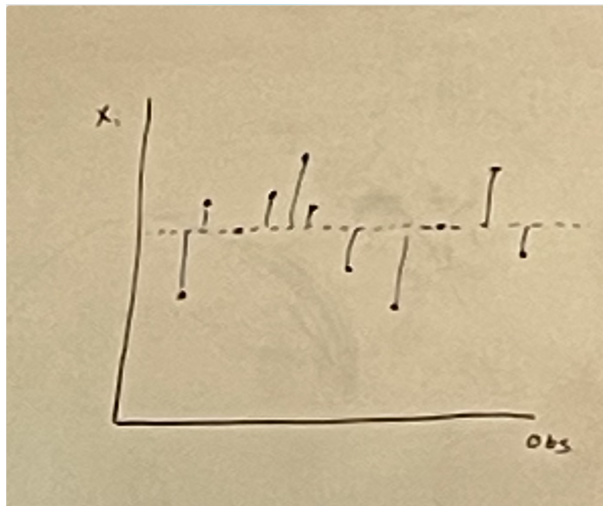


Figure 1: Dotted line is mean

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

Figure 1 shows all the  $x_i$  points as the dots and the dashed line as  $\bar{x}$ . Notice that the y-axis is  $x$ , and the x-axis is just showing us different observations, which we denote at  $i$ .

We can define the mean as the constant  $\bar{x}$  that minimizes the differences between itself and all other  $x_i$  observations,

$$\bar{x} = \mathbf{argmin} \sum_i^N (x_i - \bar{x})^2$$

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

## 2.1 Deriving eqn for mean

**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  $(x - y)^2$ .

To achieve this objective, we first need a function that measures the total distance of all  $x_i$  and  $\bar{x}$ . That function is written as:

$$F(\bar{x}) = \sum_i^N (x_i - \bar{x})^2 \quad (1)$$

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_N - \bar{x})^2 \quad (2)$$

If we're wanting to *minimize* (getting back to the argmin) the distance from  $x_i$  to  $\bar{x}$ , we can take a derivative of Equation 2 and set it equal to zero:

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

We can simplify by dividing both sides by  $-2$ , then group the  $\bar{x}$  into one term,

$$x_1 + x_2 + x_3 + \dots x_n - N\bar{x} = 0 \quad (3)$$

$$x_1 + x_2 + x_3 + \dots x_n = N\bar{x} \quad (4)$$

simplify the left-hand side of 4 by changing it back to summation notation

$$\sum_i^N x_i = N\bar{x} \implies \quad (5)$$

$$\bar{x} = \frac{1}{N} \sum_i^N x_i \quad (6)$$

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.

## 2.2 Problem set hint

The hardest problem on the problem set is a challenge problem 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  $F(x_i)$ ,

$$\text{distance} = (y_i - F(x_i))^2. \tag{7}$$

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.