Math for Data Science

Calculus - Partial Derivatives

Joanna Bieri DATA100

Important Information

- Email: joanna_bieri@redlands.edu
- Office Hours take place in Duke 209 unless otherwise noted – Office Hours Schedule

Today's Goals:

- Extend our understanding of derivatives to functions of more variables
- Calculate partial derivatives
- More Optimization!

(Review) Derivative Definition

$$\frac{dy}{dx} = \lim_{dx \to 0} \frac{f(x + dx) - f(x)}{dx}$$

(Review) First Derivative Test

- The function has a **local maximum or minimum whenever the derivative equals zero**:

$$\frac{dy}{dx} = 0$$

we solve this for all of the x-values where this is true, x^*.

- The points $(x^*, y(x^*))$ where the derivative equations zero are called "stationary points" or "critical points"
- We can tell if a critical point is a **local max** by the **derivative going from increasing to decreasing**, positive to negative, as we cross the point.
- We can tell if a critical point is a **local min** by the **derivative going from decreasing to increasing**, negative to positive, as we cross the point.

(Review) Second Derivative Test

The second derivative tells us about the curvature of the function

The notation for the second derivative:

$$\frac{d^2y}{dx^2}$$

- If the second derivative is positive then the function is concave up and our critical point is a local min
- If the second derivative is negative then the function is concave down and our critical point is a local max

(Review) Global Maximum and Minimum

Check y-values of all of the critical points but **also check the end points** or take the limit if there are no clear endpoints.

Functions of more than one variable

Most of the time in Data Science you will have more than one variable for the data you are interested in. In fact, you often have 10's 100's or 1000's of variables that you are dealing with. We will start our discussion with functions of two variables.

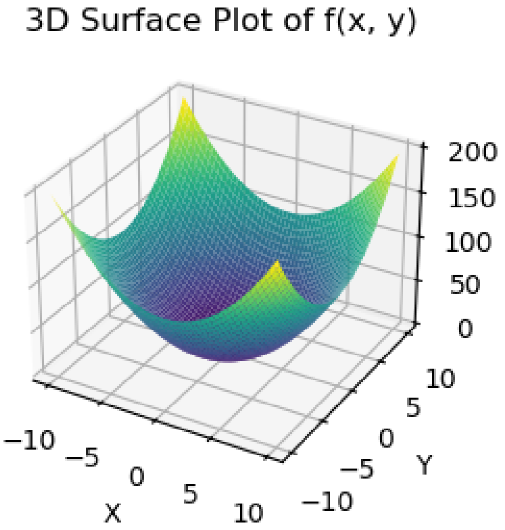
Imagine that now instead of a function that given x returns y which is a height above the x-axis we now have a function of two variables:

$$z = f(x, y)$$

In this function we give x and y values and it returns a z value which is a height above the xy-plane. Here is a plot of one such function:

$$z = x^2 + y^2$$

Functions of more than one variable



Questions

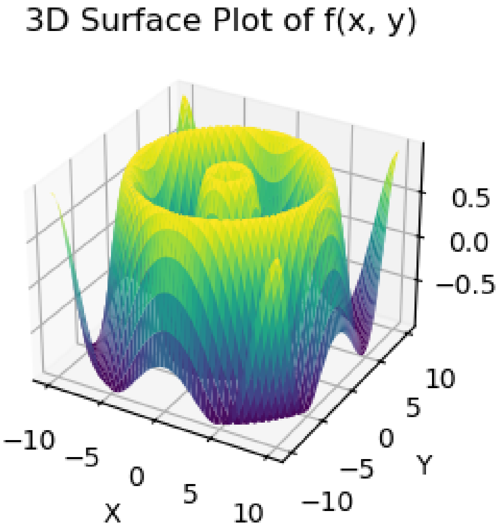
- Does this function have a slope?
- Does this function have curvature?
- Does this function have a maximum or minimum?
- Could we take the limit of this function?

Functions of more than one variable

Let's look at another more complicated function.

$$z = \sin(\sqrt{x^2 + y^2})$$

Functions of more than one variable



Questions

- If we can visualize a function of two variables as a surface, how do we visualize a function of three or more variables?
- Can we do math with functions of three or more variables?

Even more variables

For three variables $\omega = f(x, y, z)$ we could use something called level surfaces, which are "slices" of the function where the ω value is held constant. Kind of like a $3D$ topographic map.

For more than three variables it is impossible to visualize using a graph, we just run out of axes in 3D space!

BUT we can absolutely do math no mater how many variables we have! This is the power of abstract thinking.

Functions of Two Variables

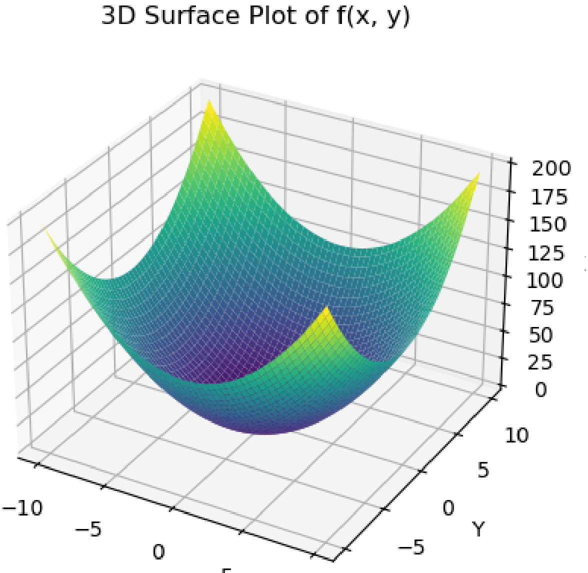
We will develop our ideas for functions of two variables, since they are easy to visualize, but remember these ideas will work no mater how many variables we have.

When we looked at the function

$$z = x^2 + y^2$$

we could see that it has slope and curvature and even a minimum!

Partial Derivatives - walking on the graph



Partial Derivatives - walking on the graph

- If I started walking on the $y = -10$ side (closest to us) and walked in a straight line along $x = 0$ what would my walk be like right at that moment?

I would be walking downhill!

- Could I say that the slope of a tangent line at that point in only the y direction would be negative?

Sure, works for me!

- What does this sound like?

Um... a derivative?

Partial Derivative in the y direction (and x direction)

We can define this idea EXACTLY like we did with the ordinary derivative. We just keep y as a constant and think about tangent lines going just in the y direction.

$$\frac{\partial z}{\partial y} = \lim_{dy \to 0} \frac{f(x, y + dy) - f(x, y)}{dy}$$

But couldn't we do the same thing in the x direction?

$$\frac{\partial z}{\partial x} = \lim_{dx \to 0} \frac{f(x + dx, y) - f(x, y)}{dx}$$

Video:

{{https://youtu.be/V-rchWU92ac}}

Sympy can do these too!

```
x,y = sp.symbols('x,y')
z = x**2+y**2

sp.diff(z,y)
sp.diff(z,x)

sp.diff(z,y,y)
sp.diff(z,x,x)
sp.diff(z,x,y)
```

Interpreting the Partial Derivative

Just like with the ordinary derivative we are talking about slope. So we can see where the slope is negative or positive. But we can also look for maximums and minimums!

Question

- What was our process for finding critical points in functions of one variable?
- What might our process be now?

Critical Points

In this case we still want to look for where the slope is equal to zero, but now we have two slopes to consider:

$$\frac{\partial z}{\partial y} = 0$$

and

$$\frac{\partial z}{\partial x} = 0$$

Critical Points

in our example above

$$\frac{\partial z}{\partial y} = 2y = 0 \rightarrow y = 0$$

$$\frac{\partial z}{\partial x} = 2x = 0 \rightarrow x = 0$$

so we have one critical point $(0, 0, z(0, 0)) = (0, 0, 0)$. Is this a max or a min?

Critical Points

Is this a max or a min?

We can look at the second derivatives!

Critical Points

Here is the notation for these second derivatives - which still talk about curvature:

$$\frac{\partial^2 z}{\partial y^2}$$

$$\frac{\partial^2 z}{\partial x^2}$$

$$\frac{\partial^2 z}{\partial y \partial x} = \frac{\partial^2 z}{\partial x \partial y}$$

We have to consider all three of the partial derivatives to be sure we have a max or min.

Critical Points

Is this a max or a min?

Here is the formula:

The determinant of the Hessian matrix (D) is calculated as:

$$D = \frac{\partial^2 z}{\partial y^2}\frac{\partial^2 z}{\partial x^2} - \left(\frac{\partial^2 z}{\partial y \partial x}\right)^2$$

- $D > 0$ and $\frac{\partial^2 z}{\partial x^2} > 0$: Local minimum
- $D > 0$ and $\frac{\partial^2 z}{\partial x^2} < 0$: Local maximum
- $D < 0$: Saddle point
- $D = 0$: Test is inconclusive

Critical Points

Is this a max or a min?

In our example:

$$D = (2)(2) - 0 > 0$$

So we have a local minimum. This makes sense because we see that BOTH the second derivatives in x and y indicate that the function is concave up (valley) and the mixed partial derivative is not going to overcome this.

FYI a non-zero mixed partial derivative indicates a kind of "twisting - curvature" in the surface.

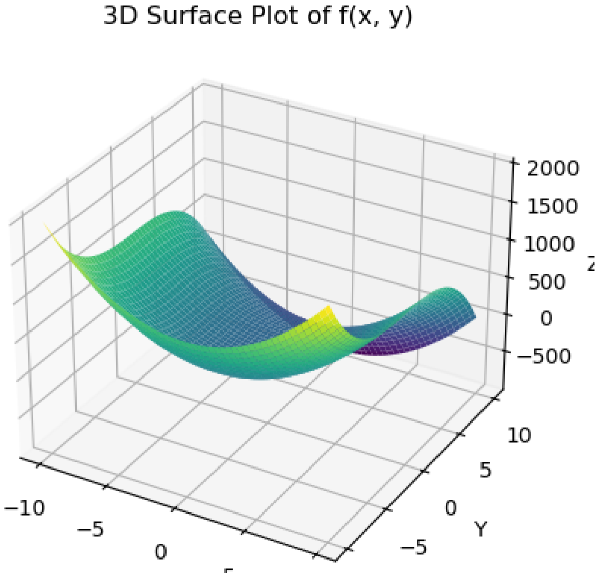
Derivatives - Critical Points - Another Example

Consider the function

$$z = 10x^2 - y^3$$

1. Plot the function
2. Take the derivatives
3. Interpret the derivatives and critical points.

Derivatives - Critical Points - Another Example



Derivatives - Critical Points - Another Example

`The function:`

$10x^2 - y^3$

`dz/dx`

$20x$

`dz/dy`

$-3y^2$

Derivatives - Critical Points - Another Example

`d2z/dx2`

20

`d2z/dxdy`

0

`d2x/xy2`

$-6y$

Derivatives - Critical Points - Another Example

To have a critical point I would need BOTH $2x = 0$ and $-3y^2 = 0$. This happens when $x = y = 0$ or at the point $(0, 0, 0)$.

I have to decide if this is a maximum or minimum - look at the second derivatives **and plug in the critical points**! The second derivative in x is positive, but the other two are zero. This means that $D = 0$ so we have a **Saddle Point**

Saddle Points

Saddle Point Saddle points are points where the function is s-shaped in one direction and u-shaped in the other but the slope right in the middle is zero. Look back at the graph. In the x-direction we do have a minimum, but in the y-direction it is the flat part of the s-shape.

To have a verified maximum or minimum we would need BOTH directions to be conclusive.

Application - Remember Least Squares?

Remember when we were doing linear regression and we said we needed to minimize the mean squared error... here is a reminder:

The goal is to **minimize** the **sum** of **square distances** between the line $\hat{y} = \beta_0 + \beta_1 x$ and the data point values y. Lets look at this is parts:

Application - Remember Least Squares?

1. What is the distance between y and $\hat{y}$ for each point in the data?

$$y_i - \hat{y}_i = y_i - (\beta_0 + \beta_1 x_i)$$

This is also called the *residual* or *error* for point i in our data.

2. What is the square distance?

$$(y_i - (\beta_0 + \beta_1 x_i))^2$$

We just square the residual or error for point i in our data.

Application - Remember Least Squares?

3. How do we add these up?

Using the summation notation:

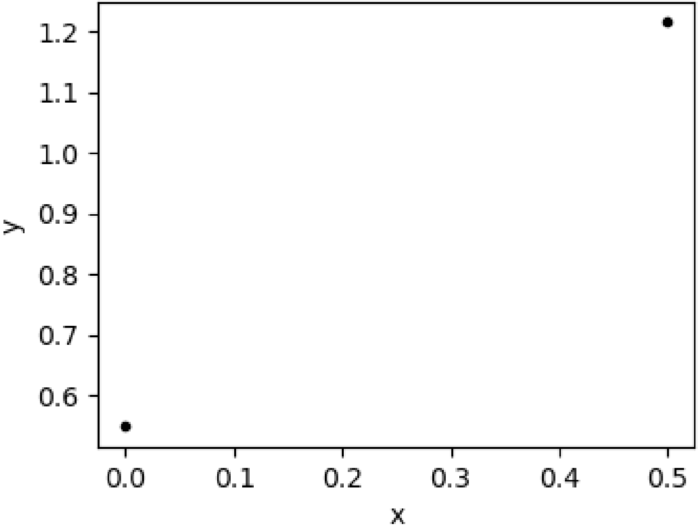
$$\sum_i (y_i - (\beta_0 + \beta_1 x_i))^2$$

Now we are adding up the square error for all of the points in the data.

4. How can we minimize this?

We need to choose β_0 and β_1 that are a minimum for this function. Lets look at the function for our example data.

Imagine we just have two data points:



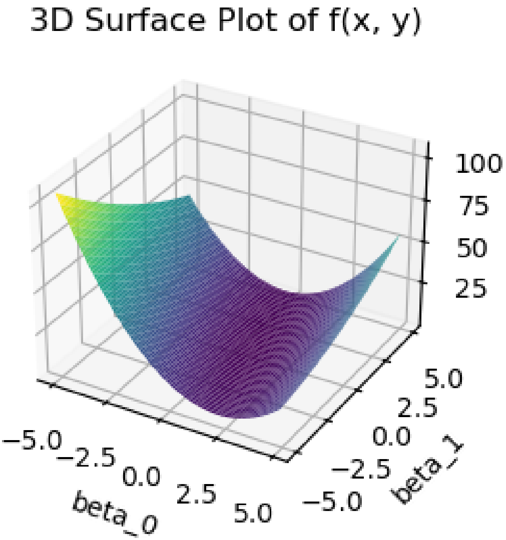
Application - Remember Least Squares?

Get the symbolic representation - Sympy

$$(0.548813503927325 - b_0)^2 + 1.47668519961446\,(-0.822917010033751 b_0 - 0.411458505016876 b_1 + 1)^2$$

This is a function of two variables and we are looking for a minimum, $f = f(\beta_0, \beta_1)$.
We will plot this function and then find the critical points!

Application - Remember Least Squares?



Does the least squares function have a minimum?

We can check - but YES - this is a special kind of function **a quadratic** and it will have a max or min. Let's do the math!

Application - Remember Least Squares?

`df/db0`

$4.0b_0 + 1.0b_1 - 3.52800574059949$

`df/db1`

$1.0b_0 + 0.5b_1 - 1.21518936637242$

We need two equations to be equal to zero

$$4.0b0 + 1.0b1 - 3.52800574059949 = 0$$

$$1.0b0 + 0.5b1 - 1.21518936637242 = 0$$

Lets use sympy to solve this.

```
{b0: 0.548813503927325, b1: 1.33275172489019}
```

Application - Remember Least Squares?

We found a critical point! $\beta_0 = 0.548813503927325$ and $\beta_1 = 1.33275172489019$. Is this a max, min, or saddle?

Application - Remember Least Squares?

`d2f/db02`

4.0

`d2f/db12`

0.5

`d2f/db0db1`

1.0

Application - Remember Least Squares?

$$D = (4)(0.5) - 1 = 1 > 0$$

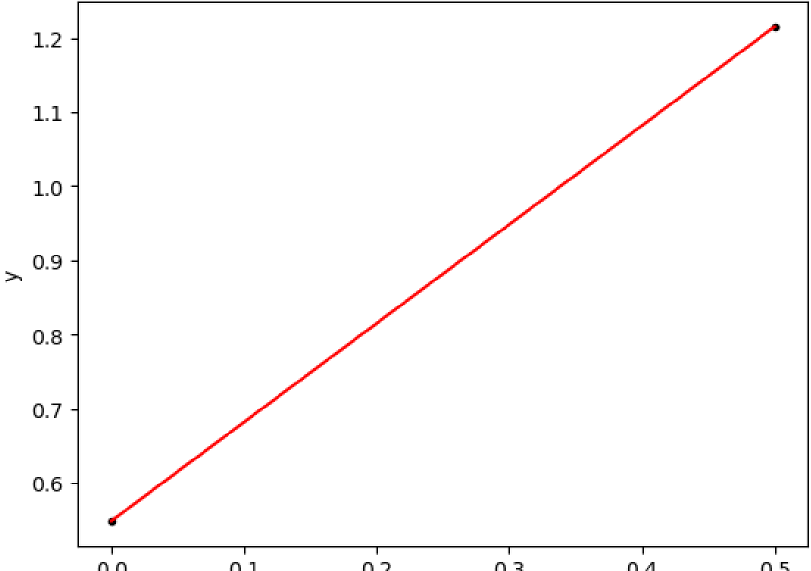
$$\frac{\partial^2 z}{\partial x^2} > 0$$

So this is a minimum. These values of β minimize the square error function and they should be the line of best fit for our data:

$$y = \beta_0 + \beta_1 x = 0.548813503927325 + 1.33275172489019x$$

Lets look at the graph of our points and this line

Application - Remember Least Squares?

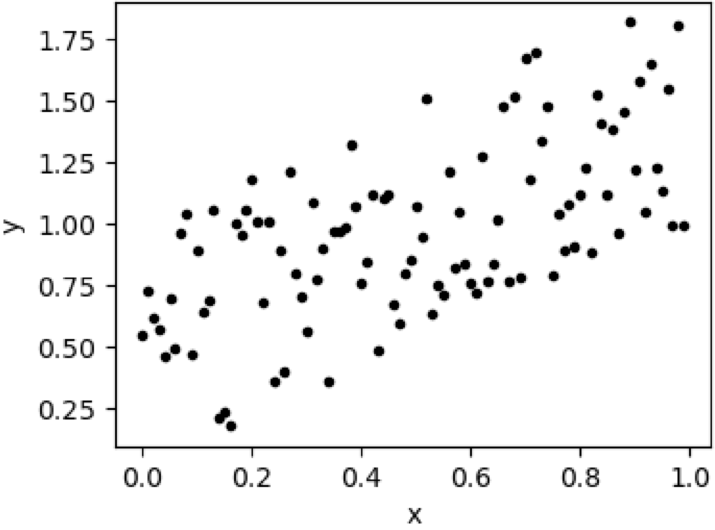


Application - Remember Least Squares?

IT WORKED!!!

Now we can do this same thing for even more points!

Application - Remember Least Squares?



Application - Remember Least Squares?

$$(0.548813503927325 - b_0)^2 +$$
$$(-b_0 - 0.99b_1 + 0.994695476192547)^2 +$$
$$(-b_0 - 0.97b_1 + 0.990107546187494)^2 +$$
$$(-b_0 - 0.87b_1 + 0.963940510758442)^2 +$$
$$(-b_0 - 0.82b_1 + 0.884147496348784)^2 +$$
$$(-b_0 - 0.79b_1 + 0.908727718954244)^2 +$$
$$(-b_0 - 0.77b_1 + 0.890196561213169)^2 +$$
$$(-b_0 - 0.75b_1 + 0.789187792254321)^2 +$$
$$(-b_0 - 0.69b_1 + 0.786098407893963)^2 +$$
$$(-b_0 - 0.67b_1 + 0.767101275793061)^2 +$$
$$(-b_0 - 0.64b_1 + 0.836582361680054)^2 +$$
$$(-b_0 - 0.63b_1 + 0.768182951348614)^2 +$$
$$(-b_0 - 0.61b_1 + 0.720375141164305)^2 +$$
$$(-b_0 - 0.6b_1 + 0.75896958364552)^2 +$$

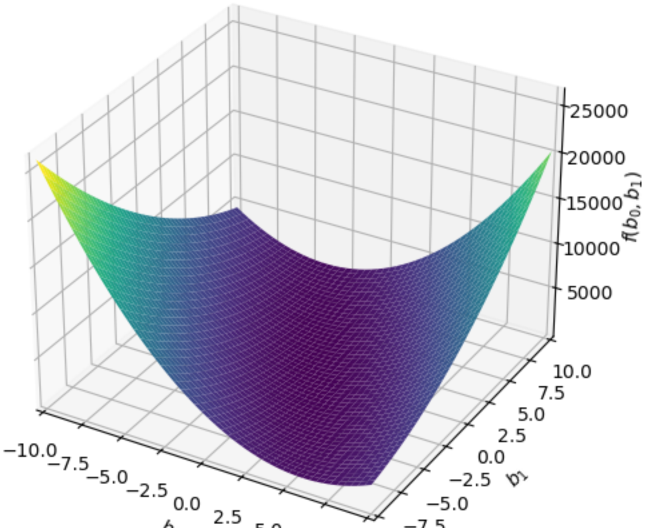
Application - Remember Least Squares?

This looks scary, but look closer! This is just a function of two variables! We could do the same thing:

1. Take partial derivatives.
2. Find critical points.
3. Check if there is a max or min.

And if we look at a plot it does have the same quadratic shape!

Application - Remember Least Squares?



Application - Remember Least Squares?

This is where we got those formulas that are used for general least squares!

$$\beta_1 = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n\sum_i x_i^2 - (\sum_i x_i)^2)}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

where $\bar{x}$ and $\bar{y}$ are the averages of the x_i and y_i data points respectively.

You Try

Find all the critical points and classify them as maximum, minimum, or saddle using the partial derivative.

- Number 1
$$z = 5e^{-(x^2+2*y^2)}$$
- Number 2
$$z = -x^4 + 4x^2 - y^2(x - 1.5)$$
  for this one I changed my x and y range in the graph a few times $[-2, 2]$,$[-5, 5]$,$[-10, 10]$, this gave a much better view of the function!

You Try

1. Copy and paste the 3d plotting code and use it to plot the function.
    - You might need to change the range of x and y to get the best view of the function
2. Enter the function using sympy
3. Take all the derivatives
4. Find any critical points
    - Use sp.solve() to find where the first derivative equations are zero
    - Ignore any answers with imaginary numbers
    - It is okay to get more than one critical point!

You Try

5. Classify them as global/local/max/min using D

   - If you have more than one critical point, you will have to calculate D separately for each one
   - Remember to plug the critical point into your derivatives

```
dxx = diff(f,x,x) # Take the derivaitve
dxx.subs({x:0,y:0}) # Plug in x=0 AND y=0
```

   - Then plug those values into D

6. Find the value of your function at the critical point - sub in the critical point

```
z.subs({x:0,y:0})
```

7. Interpret your results

You can see my solutions in the lecture notes