Basics of Derivatives

Aims

- Understand derivatives
- Understand partial derivatives
- Derivatives of common functions
- Rules of derivatives

Understand derivatives

Consider a funcation (parabola)

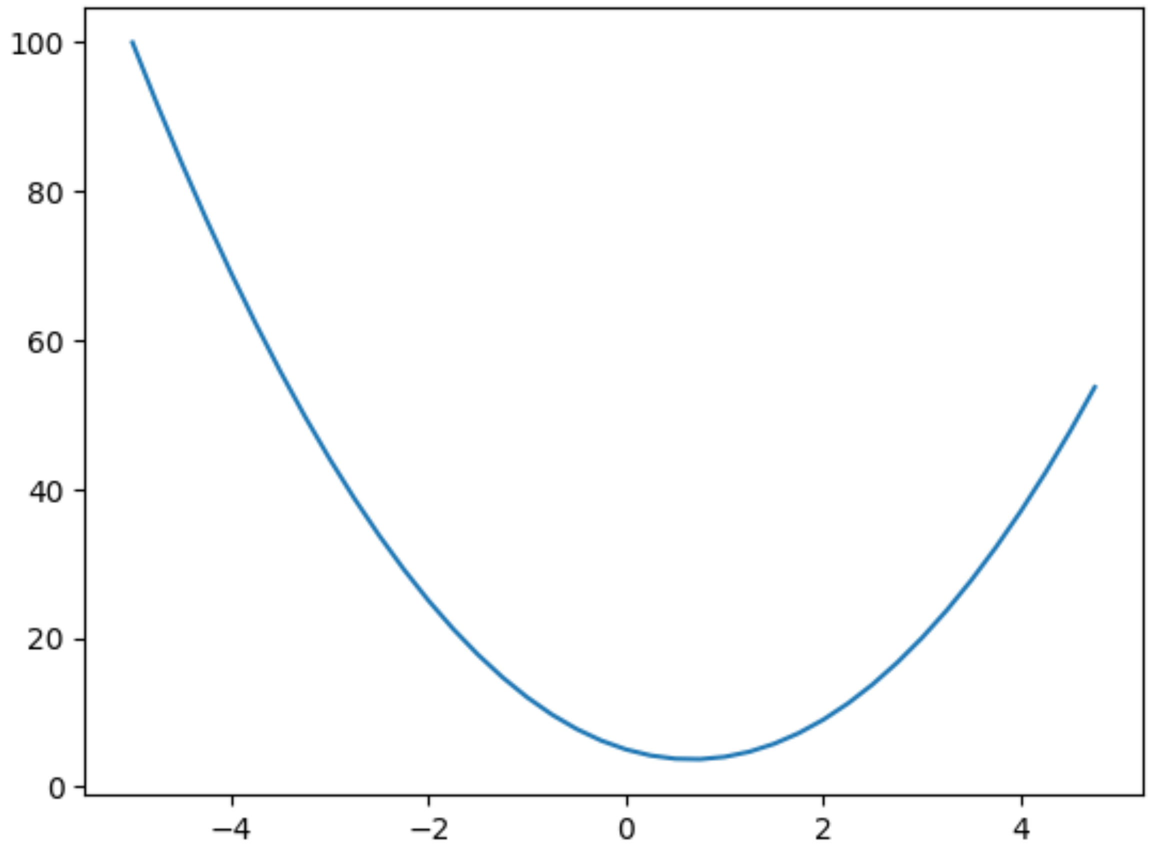
$$f(x) = 3x^2 - 4x + 5$$

In [4]:
```
import math
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [5]:
```
def f(x):
    return 3 * (x**2) - 4*x + 5
```

In [6]:
```
xs = np.arange(-5, 5, 0.25)
ys = f(xs)
plt.plot(xs, ys)
```

Out[6]:
```
[<matplotlib.lines.Line2D at 0x10dca2340>]
```



What is the derivative of f(x) at 3 ?

In [7]:
```
def df(x):
    h = 0.0000001
    return (f(x + h) - f(x))/h
```

In [8]:
```
df(3)
```

Out[8]: 14.000000305713911

What is the derivative of f(x) at -3 ?

In [9]:
```
df(-3)
```

Out[9]: -21.999999688659955

What is the derivative of f(x) at 2/3 ?

In [10]:
```
df(2.0/3.0)
```

Out[10]: 2.9753977059954195e-07

In [11]:
```
x = np.arange(-5,6,1)
dfxs = df(x)
```

In [12]:
```
from tabulate import tabulate
table = list(zip(x, dfxs))
print(tabulate(table, headers=['x', 'df(x)/dx'], tablefmt='grid'))
```

```
+-----+------------+
|   x |   df(x)/dx |
+=====+============+
|  -5 |        -34 |
+-----+------------+
|  -4 |        -28 |
+-----+------------+
|  -3 |        -22 |
+-----+------------+
|  -2 |        -16 |
+-----+------------+
|  -1 |        -10 |
+-----+------------+
|   0 |         -4 |
+-----+------------+
|   1 |          2 |
+-----+------------+
|   2 |          8 |
+-----+------------+
|   3 |         14 |
+-----+------------+
|   4 |         20 |
+-----+------------+
|   5 |         26 |
+-----+------------+
```

- When x is -3, if we increase x by a small value, f(x) will decrease (with a slope of -22).
- When x is 3, if we increase x by a small value, f(x) will increase (with a slope of 14).
- When x is 2/3, a slight increase in x will not change f(x).
- The derivative of a function helps us understand the impact of a variable on the function at any given point.

Derivative of a Multi-Variable Function (Partial Derivatives)

Consider the function: $f(x, y, z) = (x \cdot y) + z$

When a function depends on more than one variable, we use partial derivatives to understand the impact of any variable on the function. While calculating partial differentiation of $f(x, y, z)$ with respect to x, we treat y and z as constants.

In [13]:
```
def f_xyz(x, y, z):
    return x*y + z

def df_xyz(x, y, z, wrt):
    h = 0.001
    xyz = f_xyz(x,y,z)
    if wrt == 0:
        xyz_h = f_xyz(x+h, y, z)
    elif wrt == 1:
        xyz_h = f_xyz(x, y+h, z)
    else:
        xyz_h = f_xyz(x, y, z+h)
    return (xyz_h - xyz)/h
```

In [16]:
```
# Derivate of f_xyz(2, -3, 10) w.r.t x
print("Derivate of f_xyz(2, -3, 10) w.r.t x " + str(df_xyz(2, -3, 10,
print("Derivate of f_xyz(2, -3, 10) w.r.t y " + str(df_xyz(2, -3, 10,
print("Derivate of f_xyz(2, -3, 10) w.r.t z " + str(df_xyz(2, -3, 10,
```

```
Derivate of f_xyz(2, -3, 10) w.r.t x -3.0000000000001137
Derivate of f_xyz(2, -3, 10) w.r.t y 1.9999999999997797
Derivate of f_xyz(2, -3, 10) w.r.t z 0.9999999999994458
```

- Partial derivatives measure how a function changes with respect to one variable while keeping others constant
- For f(x,y,z) = xy + z:
  - ∂f/∂x = y (treating y,z as constants)
  - ∂f/∂y = x (treating x,z as constants)
  - ∂f/∂z = 1 (treating x,y as constants)
- Critical for understanding how each variable independently affects the output

Understanding Gradients

What is a Gradient?

A gradient is a vector of partial derivatives that tells us how a function changes with respect to each of its variables. The gradient points in the direction of steepest increase.

Simple Example

Let's take a function $f(x, y, z)$ and examine its gradient at point (2, -3, 10):

```
∇f = [-3, 2, 1]
```

What Does This Tell Us?

1. x-direction (∂f/∂x = -3):
   - Negative value means f decreases as x increases
   - To increase f, we should decrease x
   - The magnitude (3) tells us this effect is relatively strong
2. y-direction (∂f/∂y = 2):
   - Positive value means f increases as y increases
   - To increase f, we should increase y
   - The magnitude (2) indicates moderate influence
3. z-direction (∂f/∂z = 1):
   - Positive value means f increases as z increases
   - To increase f, we should increase z
   - The magnitude (1) suggests this has the smallest effect

Practical Application

- The gradient tells us how to adjust each variable to increase the function's value
- Moving in the opposite direction of the gradient decreases the function's value
- The magnitude tells us how strong this effect is

In Neural Networks

This concept is fundamental in training neural networks:

- The gradient tells us how to adjust each weight to reduce error
- We move weights in the opposite direction of the gradient (gradient descent)
- Larger gradients mean bigger adjustments, smaller gradients mean smaller

adjustments

Derivative of Common Functions

Function Type	Function	Derivative
Constant	c	0
Linear	x	1
Square	x^2	$2x$
Square Root	$\sqrt{x}$	$1/(2\sqrt{x})$
Exponential	e^x	e^x
Natural Log	$ln(x)$	$1/x$
Log (base a)	$log_a(x)$	$1/(x\ln(a))$
Sine	$sin(x)$	$cos(x)$
Cosine	$cos(x)$	$-sin(x)$
Tangent	$tan(x)$	$sec^2(x)$
Inverse Sine	$sin^{-1}(x)$	$1/\sqrt{1-x^2}$
Inverse Cosine	$cos^{-1}(x)$	$-1/\sqrt{1-x^2}$
Inverse Tangent	$tan^{-1}(x)$	$1/(1+x^2)$

Rules of Derivatives

Rule Name	Function	Derivative
Multiplication by a constant	$c \cdot f(x)$	$c \cdot \frac{d}{dx} f(x)$
Power Rule	x^n	$n \cdot x^{n-1}$
Sum Rule	$f(x) + g(x)$	$f'(x) + g'(x)$
	$f(x) - g(x)$	$f'(x) - g'(x)$
Product Rule	$f(x) \cdot g(x)$	$f(x) \cdot g'(x) + f'(x) \cdot g(x)$
Quotient Rule	$\frac{f(x)}{g(x)}$	$\frac{f'(x)g(x) - f(x) \cdot g'(x)}{g(x)^2}$
Reciprocal Rule	$\frac{1}{f(x)}$	$-\frac{f'(x)}{f(x)^2}$
Chain Rule	$f(g(x))$	$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$

Understanding the Chain Rule

Basic Concept

The chain rule helps us find the derivative of a composite function - when one function is applied to the result of another function.

Example 1

Let's take a simple composite function:

```
y = (x² + 1)³
```

To find dy/dx, we can break this into two parts:

```
Let u = x² + 1    (inner function)
Then y = u³       (outer function)
```

Applying Chain Rule

The chain rule states that:

```
dy/dx = (dy/du) × (du/dx)
```

Let's solve step by step:

1. Find dy/du (derivative of outer function)

   ```
   dy/du = 3u²
   ```
2. Find du/dx (derivative of inner function)

   ```
   du/dx = 2x
   ```
3. Multiply these together

   ```
   dy/dx = 3u² × 2x
   ```
4. Substitute back $u = x^2 + 1$

   ```
   dy/dx = 3(x² + 1)² × 2x
   ```

Example 2

Let's take another simple example:

```
y = sin(x²)
```

1. Let $u = x^2$
2. Then $y = \sin(u)$

Applying chain rule:

```
dy/dx = (d/du)sin(u) × (d/dx)x²
      = cos(u) × 2x
      = cos(x²) × 2x
```

Example 3 : Chain Rule with Triple Composition

The Function

Let's take a composite function with three layers:

```
y = e^(sin(x² + 1))
```

For a concrete example, let's use:

```
a = x² + 1          (inner function)
b = sin(a)          (middle function)
y = e^b             (outer function)
```

Applying Chain Rule

The chain rule tells us:

```
dy/dx = (dy/db) × (db/da) × (da/dx)
```

Step by Step Solution:

1. Find dy/db (derivative of outer function y = e^b)

   ```
   dy/db = e^b
   ```
2. Find db/da (derivative of middle function b = sin(a))

   ```
   db/da = cos(a)
   ```
3. Find da/dx (derivative of inner function a = x^2 + 1)

   ```
   da/dx = 2x
   ```
4. Multiply all three derivatives together:

   ```
   dy/dx = e^b × cos(a) × 2x
   ```
5. Substitute back the expressions for u and v:

   ```
   dy/dx = e^(sin(a)) × cos(x² + 1) × 2x
   dy/dx = e^(sin(x² + 1)) × cos(x² + 1) × 2x
   ```

Summary

Understanding Derivatives

- A derivative measures the rate of change of a function at a specific point. Derivative tells us how function will behave with small changes in input
- For function f(x):
  - Positive derivative: function increasing
  - Negative derivative: function decreasing
  - Zero derivative function at local minimum/maximum
- Numerical approximation can be used to calculate derivatives using small h

Partial Derivatives

- Partial derivatives measure how a function changes with respect to one variable while keeping others constant
- For $f(x,y,z) = xy + z$:
  - $\partial f/\partial x = y$ (treating y,z as constants)
  - $\partial f/\partial y = x$ (treating x,z as constants)
  - $\partial f/\partial z = 1$ (treating x,y as constants)
- Critical for understanding how each variable independently affects the output

Understanding Gradients

- Gradient is a vector of partial derivatives
- Points in direction of steepest increase
- Magnitude indicates rate of change
- Critical for optimization in machine learning
- Used in gradient descent to minimize loss functions

Chain Rule

- Chain rule is crucial for neural networks and backpropagation
- For composite functions, multiply derivatives in sequence
- Examples progress from simple ($y = (x^2 + 1)^3$) to complex ($y = e$^($\sin(x^2 + 1)$))
- Key principle: derivative of composition is product of derivatives
- Application in neural networks: computing gradients through multiple layers

In [ ]:

```

```