Derivative Review

- Machine learning uses derivatives in optimization problems.
- Optimization algorithms like *gradient descent* use derivates to actually decide whether to increase or decrease the weights in Neural Networks in order to increase or decrease any objective function.

Power Rule

- It helps find the derivative of a variable raised to a power
- If $f(x) = x^n$, then $\frac{\partial f(x)}{\partial x} = nx^{n-1}$

Chain Rule

- It is used to compute the derivative of composite functions
- $\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} * \frac{\partial y}{\partial x}$
- If $y = x^2$ and $x = z^2$, $\frac{\partial y}{\partial x} = 2x \text{ } and \text{ } \frac{\partial x}{\partial z} = 2z$, hence $\frac{\partial y}{\partial z} = 2x*2z$
- If we have a function such that $f(x,y) = x^4 + y^7$, the partial derivative with respect to x will be $\frac{\partial f}{\partial x} = 4x^3 + 0$. If we treat y as a constant and compute the partial derivative of the function with respect to y, we have $\frac{\partial f}{\partial y} = 0 + 7y^6$

Gradient Descent

1. Let say the cost function is

$$J_{m,b} = \frac{1}{N}\sum_{i=1}^{N}(Error_i)^2$$

2. If we are focusing on each error one at a time, then

$$\frac{\partial J}{\partial m} = 2. Error. \frac{\partial}{\partial m} Error$$

$$\frac{\partial J}{\partial b} = 2 \cdot Error. \frac{\partial}{\partial b} Error$$

3. If we calculate the gradient of error with respect to both m and b then

$$\frac{\partial}{\partial m}\text{Error} = \frac{\partial}{\partial m}(Y' - Y) \qquad \frac{\partial}{\partial b}\text{Error} = \frac{\partial}{\partial b}(Y' - Y)$$

$$\frac{\partial}{\partial m}\text{Error} = \frac{\partial}{\partial m}(mX + b - Y) \qquad \frac{\partial}{\partial b}\text{Error} = \frac{\partial}{\partial b}(mX + b - Y)$$

constants

constants

$$\frac{\partial}{\partial m}\text{Error} = \mathbf{x} \qquad \frac{\partial}{\partial b}\text{Error} = 1$$

4. If we plug the values back in the cost function and multiply it with the learning rate:

$$\frac{\partial J}{\partial m} = 2.\text{Error} * X * \text{Learning Rate}$$

Determines the direction to minimize the Error

Determines how large a step to take

$$\frac{\partial J}{\partial b} = 2.\text{Error} * \text{Learning Rate}$$

5. For the gradient descent, there are two key equations:

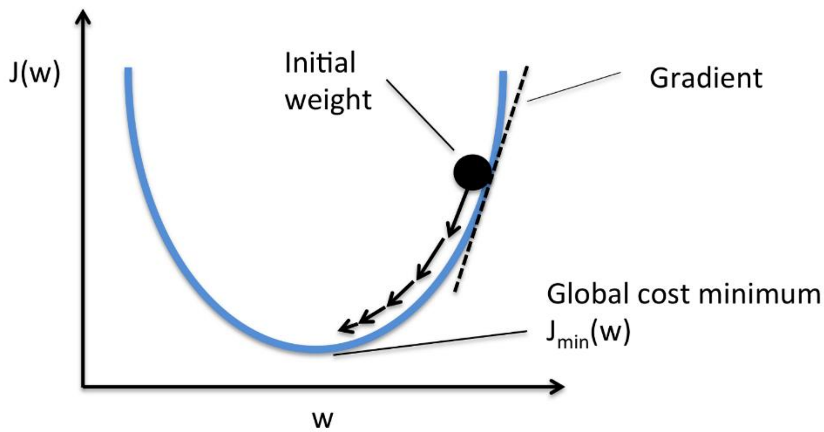
$$\frac{\partial J}{\partial m} = \text{Error} * X * \text{Learning Rate} \qquad \frac{\partial J}{\partial b} = \text{Error} * \text{Learning Rate}$$

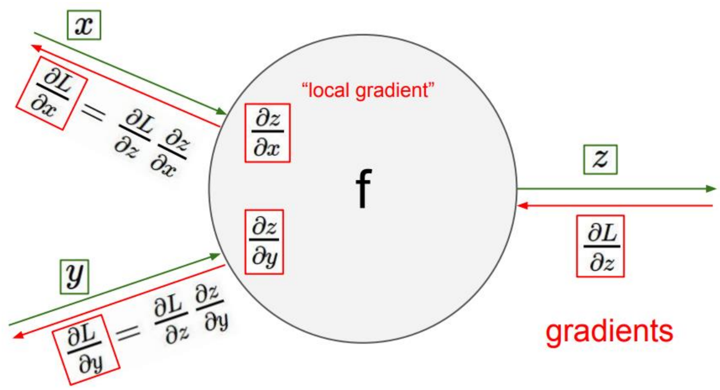
$$Since\ m = m - \delta m \qquad Since\ b = b - \delta b$$

$$m^1 = m^0 - \text{Error} * X * \text{Learning Rate} \qquad b^1 = b^0 - \text{Error} * \text{Learning Rate}$$

m^1,b^1 = next position parameters; m^o,b^o = current position parameters.

- To solve the gradient descent:
  - Iterate through the data points using the new values of m and b
  - Compute the partial derivatives.
- The new gradient indicates the slope of the cost function at a present position and the direction of update.
- The learning rate controls the magnitude of the update.


J(w)
Initial weight
Gradient
Global cost minimum
$J_{min}(w)$
w


x
$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial x}$
"local gradient"
$\frac{\partial z}{\partial x}$
f
z
$\frac{\partial z}{\partial y}$
$\frac{\partial L}{\partial z}$
y
$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial y}$
gradients

CALCULUS DERIVATIVES AND LIMITS

DERIVATIVE DEFINITION

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

BASIC PROPERTIES

$$(cf(x))' = c(f'(x))$$

$$(f(x) \pm g(x))' = f'(x) \pm g'(x)$$

$$\frac{d}{dx}(c) = 0$$

MEAN VALUE THEOREM

If f is differentiable on the interval (a, b) and continuous at the end points there exists a c in (a, b) such that

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

PRODUCT RULE

$$(f(x)g(x))' = f(x)'g(x) + f(x)g(x)'$$

QUOTIENT RULE

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

POWER RULE

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

CHAIN RULE

$$\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$$

COMMON DERIVATIVES

$$\frac{d}{dx}(x) = 1$$

$$\frac{d}{dx}(\sin x) = \cos x$$

$$\frac{d}{dx}(\cos x) = -\sin x$$

$$\frac{d}{dx}(\tan x) = \sec^2 x$$

$$\frac{d}{dx}(\sec x) = \sec x \tan x$$

$$\frac{d}{dx}(\csc x) = -\csc x \cot x$$

$$\frac{d}{dx}(\cot x) = -\csc^2 x$$

$$\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}(\cos^{-1} x) = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1+x^2}$$

$$\frac{d}{dx}(a^x) = a^x \ln(a)$$

$$\frac{d}{dx}(e^x) = e^x$$

$$\frac{d}{dx}(\ln(x)) = \frac{1}{x}, x > 0$$

$$\frac{d}{dx}(\ln|x|) = \frac{1}{x}$$

$$\frac{d}{dx}(\log_a(x)) = \frac{1}{x \ln(a)}$$

LIMIT EVALUATION METHOD – FACTOR AND CANCEL

$$\lim_{x \to -3} \frac{x^2 - x - 12}{x^2 + 3x} = \lim_{x \to -3} \frac{(x+3)(x-4)}{x(x+3)} = \lim_{x \to -3} \frac{(x-4)}{x} = \frac{7}{3}$$

L'HOPITAL'S RULE

$$\text{If } \lim_{x \to a} \frac{f(x)}{g(x)} = \frac{0}{0} \text{ or } \frac{\pm\infty}{\pm\infty} \text{ then } \lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

CHAIN RULE AND OTHER EXAMPLES

$$\frac{d}{dx}([f(x)]^n) = n[f(x)]^{n-1} f'(x)$$

$$\frac{d}{dx}(e^{f(x)}) = f'(x)e^{f(x)}$$

$$\frac{d}{dx}(\ln[f(x)]) = \frac{f'(x)}{f(x)}$$

$$\frac{d}{dx}(\sin[f(x)]) = f'(x)\cos[f(x)]$$

$$\frac{d}{dx}(\cos[f(x)]) = -f'(x)\sin[f(x)]$$

$$\frac{d}{dx}(\tan[f(x)]) = f'(x)\sec^2[f(x)]$$

$$\frac{d}{dx}(\sec[f(x)]) = f'(x)\sec[f(x)]\tan[f(x)]$$

$$\frac{d}{dx}(\tan^{-1}[f(x)]) = \frac{f'(x)}{1 + [f(x)]^2}$$

$$\frac{d}{dx}(f(x)^{g(x)}) = f(x)^{g(x)} \left(\frac{g(x)f'(x)}{f(x)} + \ln(f(x))\, g'(x)\right)$$

PROPERTIES OF LIMITS

These properties require that the limit of $f(x)$ and $g(x)$ exist

$$\lim_{x \to a}[cf(x)] = c \lim_{x \to a} f(x)$$

$$\lim_{x \to a}[f(x) \pm g(x)] = \lim_{x \to a} f(x) \pm \lim_{x \to a} g(x)$$

$$\lim_{x \to a}[f(x)g(x)] = \lim_{x \to a} f(x) \lim_{x \to a} g(x)$$

$$\lim_{x \to a}\left[\frac{f(x)}{g(x)}\right] = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)} \text{ if } \lim_{x \to a} g(x) \neq 0$$

$$\lim_{x \to a}[f(x)]^n = \left[\lim_{x \to a} f(x)\right]^n$$

LIMIT EVALUATIONS AT +-∞

$$\lim_{x \to \infty} e^x = \infty \text{ and } \lim_{x \to -\infty} e^x = 0$$

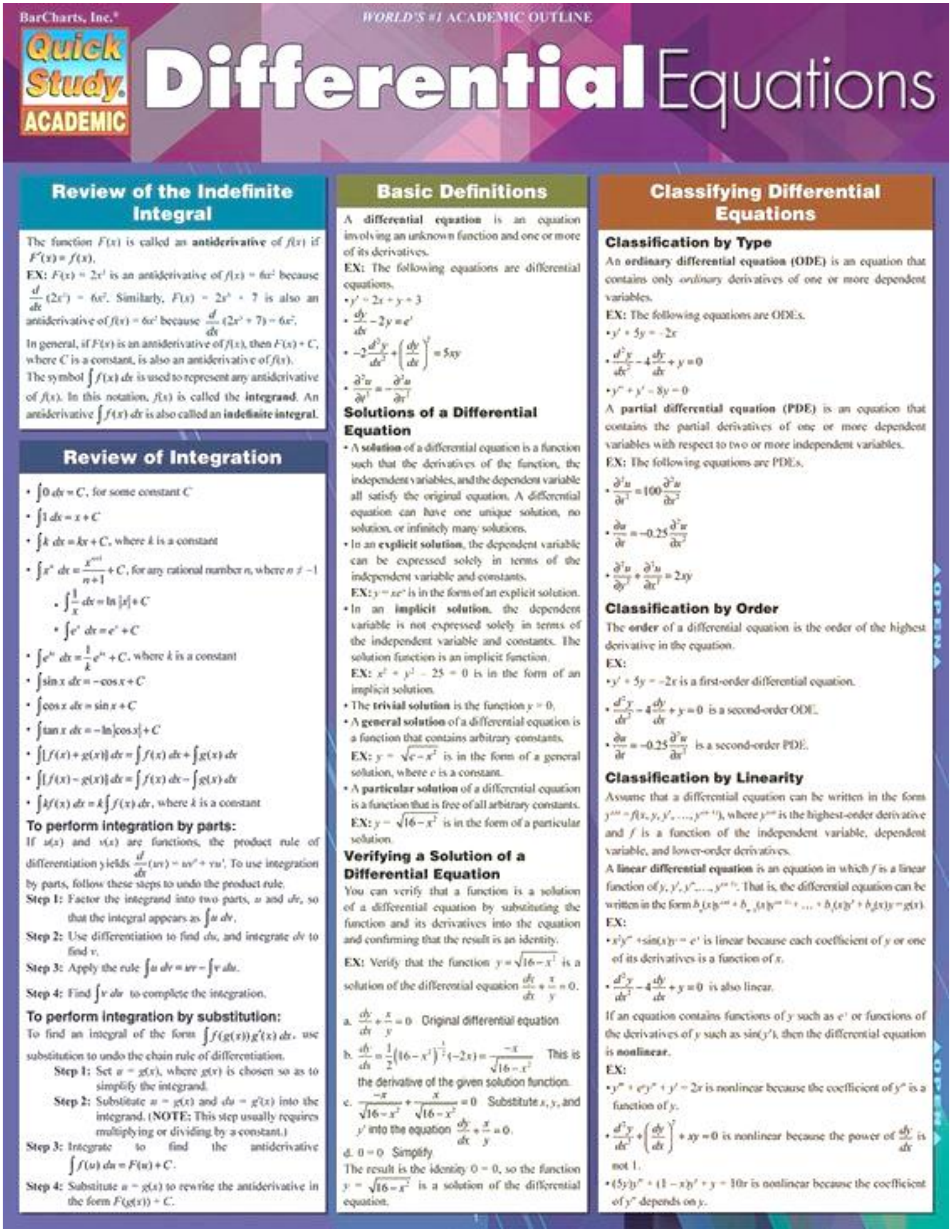
$$\lim_{x \to \infty} \ln(x) = \infty \text{ and } \lim_{x \to 0^+} \ln(x) = -\infty$$

$$\text{If } r > 0 \text{ then } \lim_{x \to \infty} \frac{c}{x^r} = 0$$

$$\text{If } r > 0 \text{ \& } x^r \text{ is real for } x < 0 \text{ then } \lim_{x \to -\infty} \frac{c}{x^r} = 0$$

$$\lim_{x \to \pm\infty} x^r = \infty \text{ for even } r$$

$$\lim_{x \to \infty} x^r = \infty \text{ \& } \lim_{x \to -\infty} x^r = -\infty \text{ for odd } r$$



BarCharts, Inc.®

WORLD'S #1 ACADEMIC OUTLINE

Quick Study. ACADEMIC

Differential Equations

Review of the Indefinite Integral

The function $F(x)$ is called an **antiderivative** of $f(x)$ if $F'(x) = f(x)$.

EX: $F(x) = 2x^3$ is an antiderivative of $f(x) = 6x^2$ because $\frac{d}{dx}(2x^3) = 6x^2$. Similarly, $F(x) = 2x^3 + 7$ is also an antiderivative of $f(x) = 6x^2$ because $\frac{d}{dx}(2x^3 + 7) = 6x^2$.

In general, if $F(x)$ is an antiderivative of $f(x)$, then $F(x) + C$, where C is a constant, is also an antiderivative of $f(x)$.

The symbol $\int f(x)\,dx$ is used to represent any antiderivative of $f(x)$. In this notation, $f(x)$ is called the **integrand**. An antiderivative $\int f(x)\,dx$ is also called an **indefinite integral**.

Review of Integration

- $\int 0\,dx = C$, for some constant C
- $\int 1\,dx = x + C$
- $\int k\,dx = kx + C$, where k is a constant
- $\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$, for any rational number n, where $n \neq -1$
  - $\int \frac{1}{x}\,dx = \ln|x| + C$
  - $\int e^x\,dx = e^x + C$
- $\int e^{kx}\,dx = \frac{1}{k}e^{kx} + C$, where k is a constant
- $\int \sin x\,dx = -\cos x + C$
- $\int \cos x\,dx = \sin x + C$
- $\int \tan x\,dx = -\ln|\cos x| + C$
- $\int [f(x) + g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx$
- $\int [f(x) - g(x)]\,dx = \int f(x)\,dx - \int g(x)\,dx$
- $\int kf(x)\,dx = k\int f(x)\,dx$, where k is a constant

To perform integration by parts:

If $u(x)$ and $v(x)$ are functions, the product rule of differentiation yields $\frac{d}{dx}(uv) = uv' + vu'$. To use integration by parts, follow these steps to undo the product rule.

Step 1: Factor the integrand into two parts, u and dv, so that the integral appears as $\int u\,dv$.

Step 2: Use differentiation to find du, and integrate dv to find v.

Step 3: Apply the rule $\int u\,dv = uv - \int v\,du$.

Step 4: Find $\int v\,du$ to complete the integration.

To perform integration by substitution:

To find an integral of the form $\int f(g(x))g'(x)\,dx$, use substitution to undo the chain rule of differentiation.

Step 1: Set $u = g(x)$, where $g(x)$ is chosen so as to simplify the integrand.

Step 2: Substitute $u = g(x)$ and $du = g'(x)$ into the integrand. (**NOTE:** This step usually requires multiplying or dividing by a constant.)

Step 3: Integrate to find the antiderivative $\int f(u)\,du = F(u) + C$.

Step 4: Substitute $u = g(x)$ to rewrite the antiderivative in the form $F(g(x)) + C$.

Basic Definitions

A **differential equation** is an equation involving an unknown function and one or more of its derivatives.

EX: The following equations are differential equations.

- $y' = 2x + y + 3$
- $\frac{dy}{dx} - 2y = e^x$
- $-2\frac{d^2y}{dx^2} + \left(\frac{dy}{dx}\right)^3 = 5xy$
- $\frac{\partial^2 u}{\partial t^2} = -\frac{\partial^2 u}{\partial x^2}$

Solutions of a Differential Equation

- A **solution** of a differential equation is a function such that the derivatives of the function, the independent variables, and the dependent variable all satisfy the original equation. A differential equation can have one unique solution, no solution, or infinitely many solutions.
- In an **explicit solution**, the dependent variable can be expressed solely in terms of the independent variable and constants.
  EX: $y = xe^x$ is in the form of an explicit solution.
- In an **implicit solution**, the dependent variable is not expressed solely in terms of the independent variable and constants. The solution function is an implicit function.
  EX: $x^2 + y^2 - 25 = 0$ is in the form of an implicit solution.
- The **trivial solution** is the function $y = 0$.
- A **general solution** of a differential equation is a function that contains arbitrary constants.
  EX: $y = \sqrt{c - x^2}$ is in the form of a general solution, where c is a constant.
- A **particular solution** of a differential equation is a function that is free of all arbitrary constants.
  EX: $y = \sqrt{16 - x^2}$ is in the form of a particular solution.

Verifying a Solution of a Differential Equation

You can verify that a function is a solution of a differential equation by substituting the function and its derivatives into the equation and confirming that the result is an identity.

EX: Verify that the function $y = \sqrt{16 - x^2}$ is a solution of the differential equation $\frac{dy}{dx} + \frac{x}{y} = 0$.

a. $\frac{dy}{dx} + \frac{x}{y} = 0$ Original differential equation

b. $\frac{dy}{dx} = \frac{1}{2}\left(16 - x^2\right)^{-\frac{1}{2}}(-2x) = \frac{-x}{\sqrt{16 - x^2}}$ This is the derivative of the given solution function.

c. $\frac{-x}{\sqrt{16 - x^2}} + \frac{x}{\sqrt{16 - x^2}} = 0$ Substitute x, y, and y' into the equation $\frac{dy}{dx} + \frac{x}{y} = 0$.

d. $0 = 0$ Simplify

The result is the identity $0 = 0$, so the function $y = \sqrt{16 - x^2}$ is a solution of the differential equation.

Classifying Differential Equations

Classification by Type

An **ordinary differential equation (ODE)** is an equation that contains only *ordinary* derivatives of one or more dependent variables.

EX: The following equations are ODEs.

- $y' + 5y = -2x$
- $\frac{d^2y}{dx^2} - 4\frac{dy}{dx} + y = 0$
- $y'' + y' - 8y = 0$

A **partial differential equation (PDE)** is an equation that contains the partial derivatives of one or more dependent variables with respect to two or more independent variables.

EX: The following equations are PDEs.

- $\frac{\partial^2 u}{\partial t^2} = 100\frac{\partial^2 u}{\partial x^2}$
- $\frac{\partial u}{\partial t} = -0.25\frac{\partial^2 u}{\partial x^2}$
- $\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial x^2} = 2xy$

Classification by Order

The **order** of a differential equation is the order of the highest derivative in the equation.

EX:

- $y' + 5y = -2x$ is a first-order differential equation.
- $\frac{d^2y}{dx^2} - 4\frac{dy}{dx} + y = 0$ is a second-order ODE.
- $\frac{\partial u}{\partial t} = -0.25\frac{\partial^2 u}{\partial x^2}$ is a second-order PDE.

Classification by Linearity

Assume that a differential equation can be written in the form $y^{(n)} = f(x, y, y', \ldots, y^{(n-1)})$, where $y^{(n)}$ is the highest-order derivative and f is a function of the independent variable, dependent variable, and lower-order derivatives.

A **linear differential equation** is an equation in which f is a linear function of $y, y', y'', \ldots, y^{(n-1)}$. That is, the differential equation can be written in the form $b_n(x)y^{(n)} + b_{n-1}(x)y^{(n-1)} + \ldots + b_1(x)y' + b_0(x)y = g(x)$.

EX:

- $x^2y'' + \sin(x)y = e^x$ is linear because each coefficient of y or one of its derivatives is a function of x.
- $\frac{d^2y}{dx^2} - 4\frac{dy}{dx} + y = 0$ is also linear.

If an equation contains functions of y such as e^y or functions of the derivatives of y such as $\sin(y')$, then the differential equation is **nonlinear**.

EX:

- $y''' + e^y y'' + y' = 2x$ is nonlinear because the coefficient of y'' is a function of y.
- $\frac{d^2y}{dx^2} + \left(\frac{dy}{dx}\right)^3 + xy = 0$ is nonlinear because the power of $\frac{dy}{dx}$ is not 1.
- $(5y)y'' + (1 - x)y' + y = 10x$ is nonlinear because the coefficient of y'' depends on y.

Sources: Toward Science, eCalc.com, BarCharts, Inc. For class use only.