

*COMP 1433: Introduction to Data Analytics &
COMP 1003: Statistical Tools and Applications*

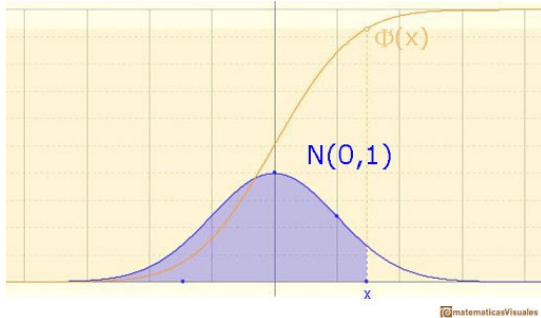
Lecture 5 – Calculus Basics

Dr. Jing Li

Department of Computing

The Hong Kong Polytechnic University

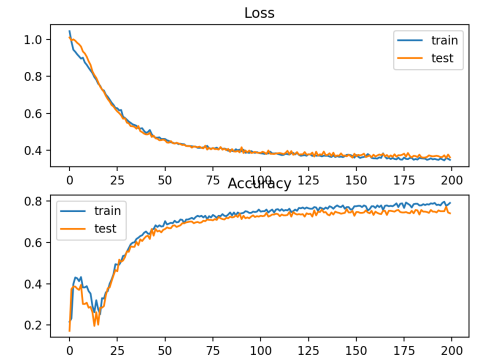
13 & 14 Feb 2023



continuous random variables



trend of time-series data



machine learning

Why we learn *Calculus*?

What are *Functions*?

- y is a *function* of x , written with $y = f(x)$:
 - Every value of x corresponds to *one and only one value* of y .
 - x is the *independent variable* while y is the *dependent variable*.
 - **EXAMPLE.** Distance traveled per hour y is a function of velocity x .



Optimization of a Function

- Given a function $f(x_1, x_2, \dots, x_n)$, where x_1, x_2, \dots, x_n are variables or parameters.
- **Optimization**: Find a set of variables x_1, x_2, \dots, x_n that maximize or minimize $f(x_1, x_2, \dots, x_n)$.
- An optimization problem in everyday life:
 - You selected 3 classes this semester. The three classes have different effects on the GPA and you want to maximize the GPA (*the value of a function*) via priorly knowing the # of hours (*function variables*) you should spend on each of the class.

A Recipe for Machine Learning

1. Given training data:

$$\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$$

Face



Face



Not a face



2. Choose each of these:

- Decision function

$$\hat{\mathbf{y}} = f_{\theta}(\mathbf{x}_i)$$

- Loss function

$$\ell(\hat{\mathbf{y}}, \mathbf{y}_i) \in \mathbb{R}$$

Examples: Linear regression, Logistic regression, Neural Network

Examples: Mean-squared error, Cross Entropy

Roadmap

- Derivatives
 - Basic Concepts of Derivatives
 - How to calculate derivatives?
 - Partial Derivatives and Gradients
 - **Application:** the training of machine learning
- Antiderivatives: Integrals
 - Basic Concepts of integrals.
 - How to calculate integrals?

Roadmap

- Derivatives

- Basic Concepts of Derivatives
- How to calculate derivatives?
- Partial Derivatives and Gradients
- **Application:** the training of machine learning

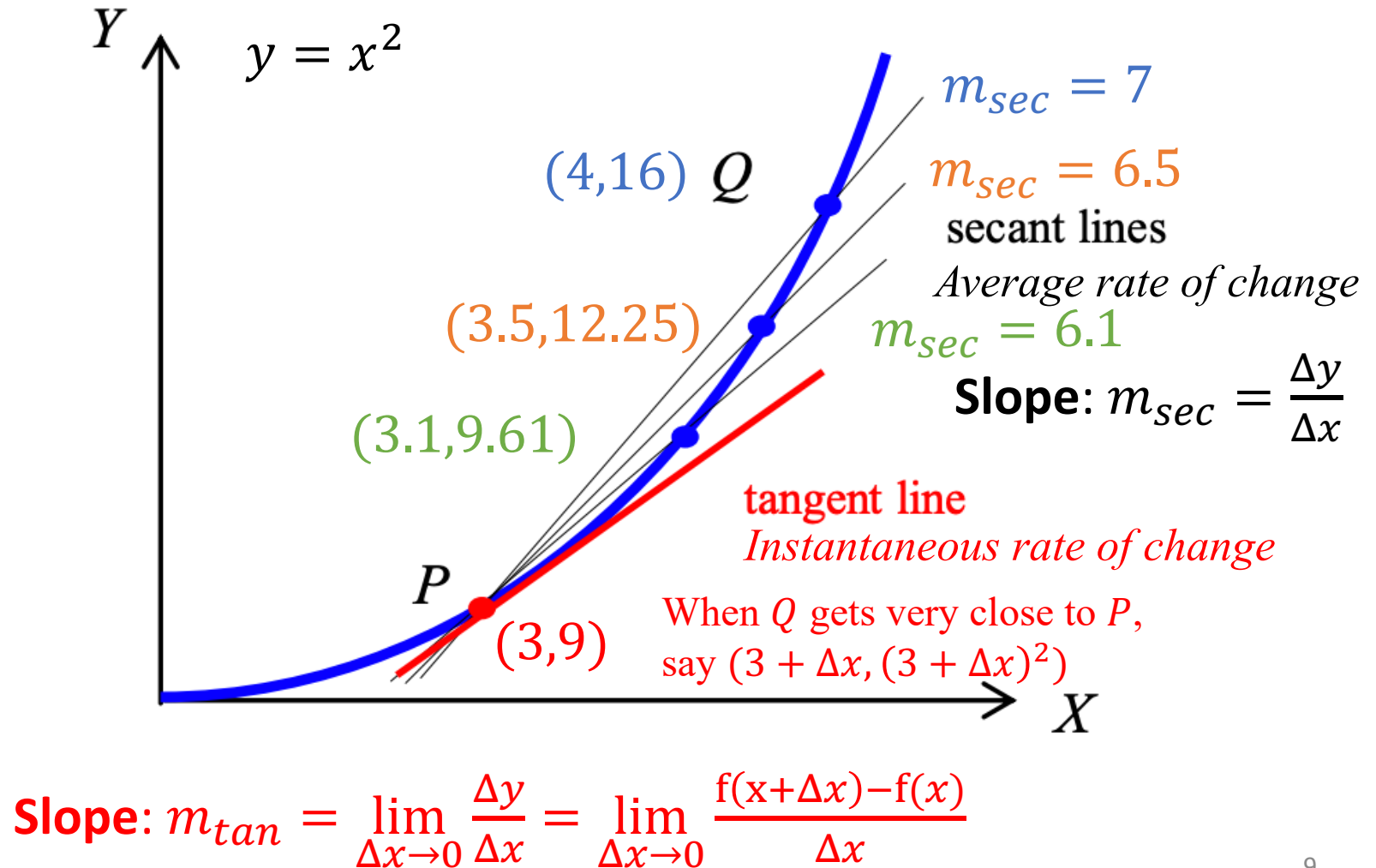
- Antiderivatives: Integrals

- Basic Concepts of integrals.
- How to calculate integrals?

Roadmap

- Derivatives
 - Basic Concepts of Derivatives
 - How to calculate derivatives?
 - Partial Derivatives and Gradients
 - **Application:** the training of machine learning
- Antiderivatives: Integrals
 - Basic Concepts of integrals.
 - How to calculate integrals?

Secant Line and Tangent Line



What are *Derivatives*?

- The *derivative* of $f(x)$ is the *slope of tangent line* (*instantaneous rate of change*) at $(x, f(x))$
 - $\frac{dy}{dx} = f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$
- The process of calculating the derivatives of a function is called *differentiation*.
- **Example:** $y = x^2$
 - $\frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x} = \lim_{\Delta x \rightarrow 0} (2x + \Delta x) = 2x$
 - Also written as: $dy = 2x dx$ (the *differential* dy in terms of the *differential* dx)
 - Used to estimate the output difference Δy in terms of a small input difference Δx

Roadmap

- Derivatives
 - Basic Concepts of Derivatives
 - How to calculate derivatives?
 - Partial Derivatives and Gradients
 - **Application:** the training of machine learning
- Antiderivatives: Integrals
 - Basic Concepts of integrals.
 - How to calculate integrals?

Some Useful Derivatives

- **Power Rule:** $\frac{d(x^p)}{dx} = px^{p-1}$
 - $\frac{d(x^3)}{dx} = 3x^2$; $\frac{d(\frac{1}{x})}{dx} = -\frac{1}{x^2}$; $\frac{dx^{\frac{1}{2}}}{dx} = \frac{1}{2}x^{-\frac{1}{2}}$
 - $\frac{dx}{dx} = 1$ (*y = x has slope 1 every where*)
- **Exponential Rule:** $\frac{d(b^x)}{dx} = b^x \ln b$
 - $\frac{d(e^x)}{dx} = e^x$
- **Logarithm Rule:** $\frac{d(\log_b x)}{dx} = \frac{1}{x \ln b}$
 - $\frac{d(\ln x)}{dx} = \frac{1}{x}$
- **Derivatives for constants:** $\frac{dC}{dx} = 0$

Properties of Derivatives

- For any constant c and any *differentiable* function $f(x)$,
$$\frac{d[cf(x)]}{dx} = c \frac{d[f(x)]}{dx}$$
 - $\frac{d[5x^3]}{dx} = 5 \cdot 3x^2 = 15x^2$
 - $\frac{d[-3e^{2x}]}{dx} = -3 \cdot \frac{d[(e^2)^x]}{dx} = -3e^{2x} \ln(e^2) = -6e^{2x}$
- For any two *differentiable* functions: $f(x)$ and $g(x)$
 - **Sum and Difference Rules.**
 - **Product Rule.**
 - **Quotient Rule.**
 - **Chain Rule.**

Properties of Derivatives (cont.)

- For any two *differentiable* functions: $f(x)$ and $g(x)$
 - **Sum and Difference Rules.**
 - $\frac{d[f(x) \pm g(x)]}{dx} = \frac{df(x)}{dx} \pm \frac{dg(x)}{dx}$
 - **Example:** $y = x^{\frac{3}{2}} - 7x^4 + 10e^{-3x} - 5$
 - $\frac{dy}{dx} = \frac{3}{2}x^{\frac{1}{2}} - 28x^3 - 30e^{-3x}$
 - **Product Rule.**
 - **Quotient Rule.**
 - **Chain Rule.**

Properties of Derivatives (cont.)

- For any two *differentiable* functions: $f(x)$ and $g(x)$
 - **Sum and Difference Rules.**
 - **Product Rule.**
 - $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$
 - **Example:** $y = x^{11}e^{6x}$
 - $\frac{dy}{dx} = 11x^{10}e^{6x} + 6e^{6x}x^{11} = (11 + 6x)x^{10}e^{6x}$
 - **Quotient Rule.**
 - **Chain Rule.**

Properties of Derivatives (cont.)

- For any two *differentiable* functions: $f(x)$ and $g(x)$
 - **Sum and Difference Rules.**
 - **Product Rule.**
 - **Quotient Rule.**
 - $\left[\frac{f(x)}{g(x)}\right]' = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$ where $g(x) \neq 0$
 - **Example:** $y = \frac{e^{4x}}{x^7 + 8}$
 - $\frac{dy}{dx} = \frac{4e^{4x}(x^7 + 8) - e^{4x}(7x^6)}{(x^7 + 8)^2} = \frac{e^{4x}(4x^7 + 32 - 7x^6)}{(x^7 + 8)^2}$
 - **Chain Rule.**

Properties of Derivatives (cont.)

- For any two *differentiable* functions: $f(x)$ and $g(x)$
 - **Sum and Difference Rules.**
 - **Product Rule.**
 - **Quotient Rule.**
 - **Chain Rule.**
 - $[f(g(x))]' = f'(g(x)) \cdot g'(x)$
 - **Example:** $y = \left(x^{\frac{2}{3}} + 2e^{-9x}\right)^6$
 - $\frac{dy}{dx} = 6 \left(x^{\frac{2}{3}} + 2e^{-9x}\right)^5 \cdot \left(\frac{2}{3}x^{-\frac{1}{3}} - 18e^{-9x}\right)$

Properties of Derivatives (cont.)

- For any two *differentiable* functions: $f(x)$ and $g(x)$
 - **Sum and Difference Rules.**
 - **Product Rule.**
 - **Quotient Rule.**
 - **Chain Rule.**
 - $[f(g(x))]' = f'(g(x)) \cdot g'(x)$
 - **Example:** $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
 - $\frac{dy}{dx} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot (-x) = -\frac{1}{\sqrt{2\pi}} x \cdot e^{-\frac{x^2}{2}}$

Roadmap

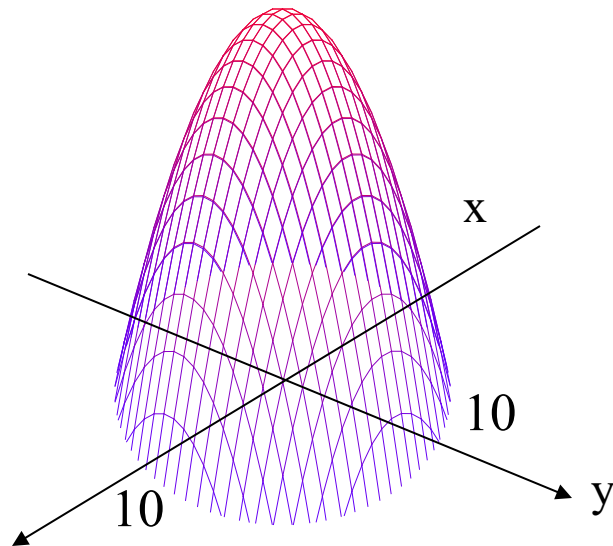
- Derivatives

- Basic Concepts of Derivatives
- How to calculate derivatives?
- Partial Derivatives and Gradients
- **Application:** the training of machine learning

- Antiderivatives: Integrals

- Basic Concepts of integrals.
- How to calculate integrals?

Partial Derivatives



$$f(x, y) = 100 - x^2 - y^2$$

$$\frac{\partial f}{\partial x} = -2x \quad \frac{\partial f}{\partial y} = -2y$$

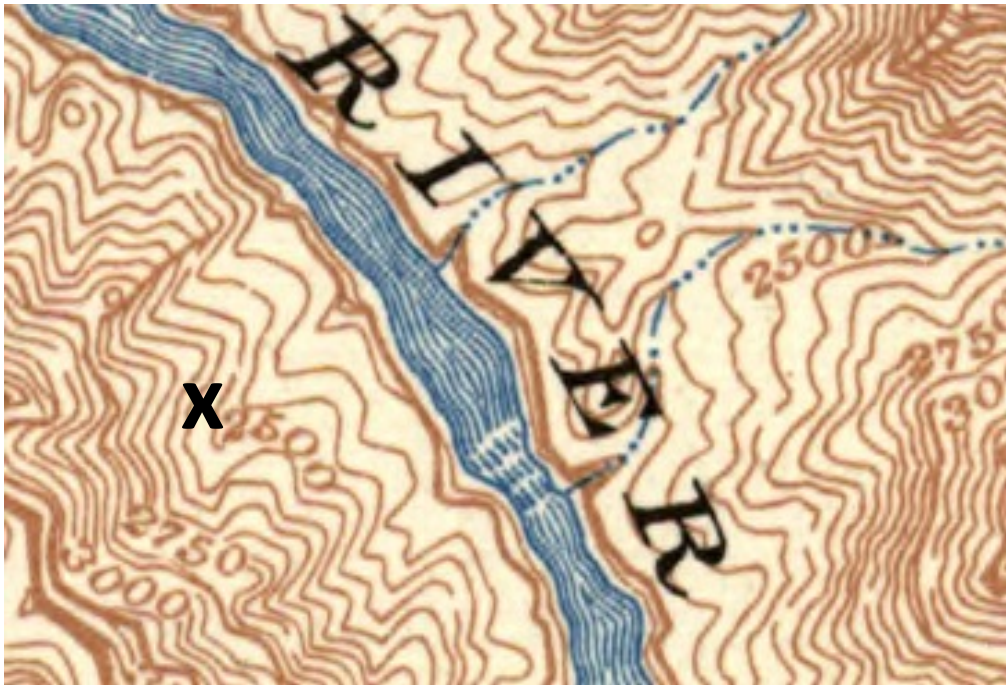
- A function may have multiple variables, e.g.,
 - $f(x, y) = x^2y$ and $g(x_1, x_2, x_3) = x_1x_2x_3$
- A **partial derivative** of a function of several variables is its derivative with respect to one of those variables, with the others held constant
- Usually denoted by:
 - $\frac{\partial f}{\partial x}$ or simply $\frac{df}{dx}$

Gradient

- Given a function $f(x_1, x_2, \dots, x_n)$ with multiple variables x_1, x_2, \dots, x_n
- Measure the partial derivatives for each variable
- A **gradient** is a *vector* holding all partial derivatives:
 - $\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$
- ∇f points in the direction of greatest rate of change (or “steepest ascent”)
- **Application:** *Gradient Descent Algorithm* for the training of most machine learning models.

Intuition of gradient descent

- How do I get to the bottom of this river canyon?

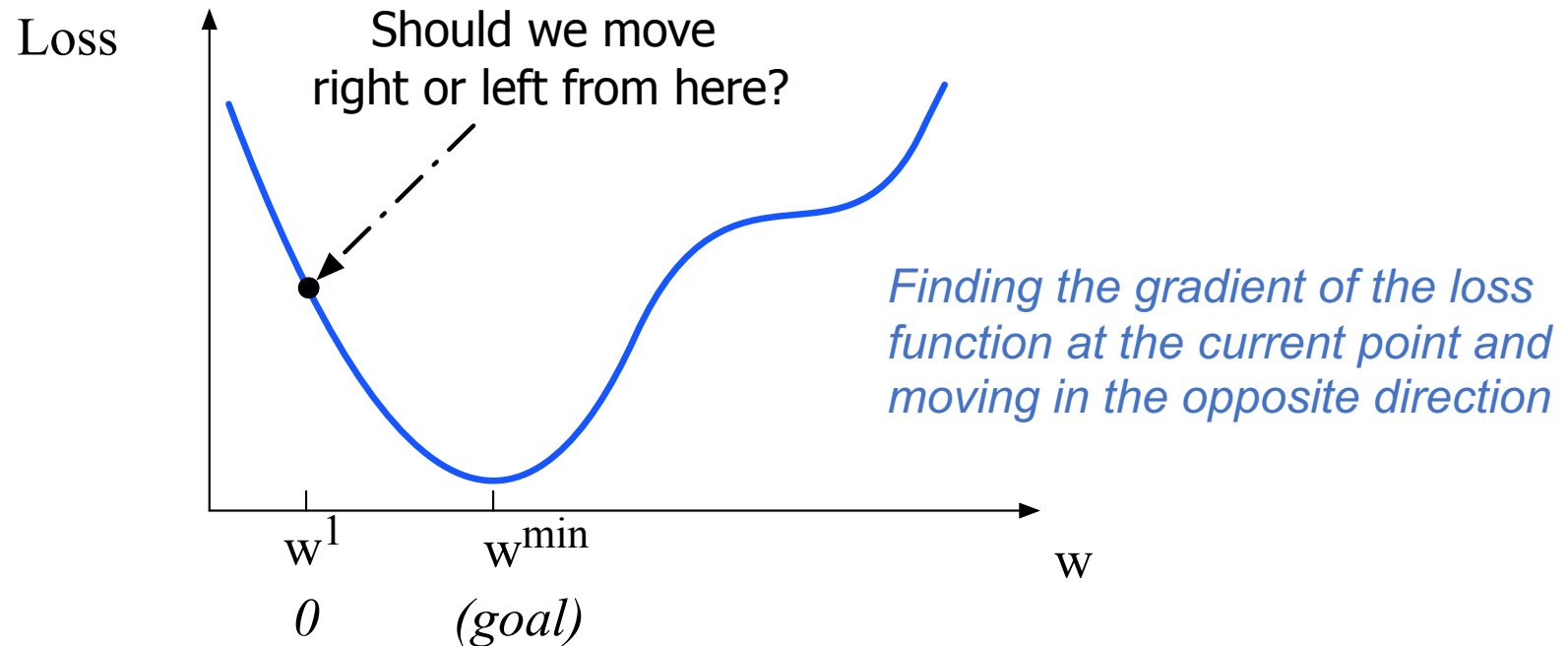


Look around me 360°
Find the direction of
steepest slope down
Go that way

Let's first visualize for a single scalar w

Q: Given current w , should we make it bigger or smaller?

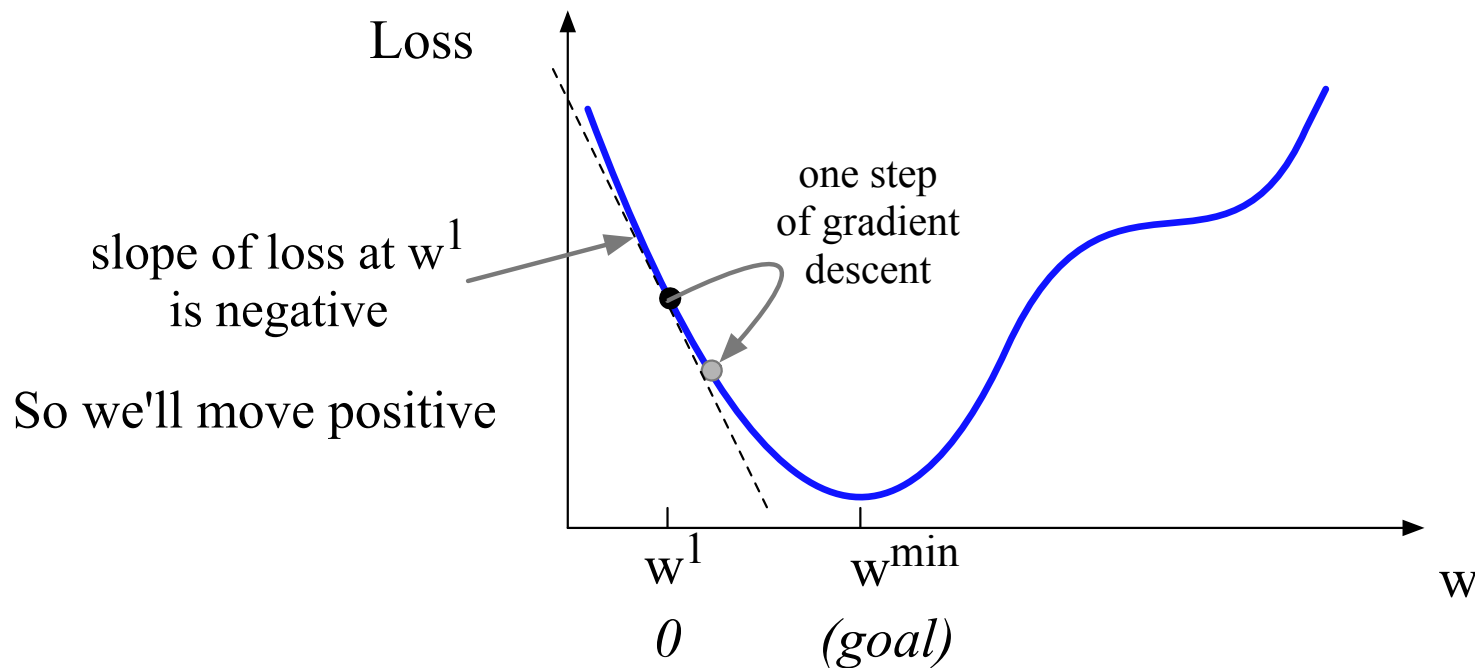
A: Move w in the reverse direction from the slope of the function



Let's first visualize for a single scalar w

Q: Given current w , should we make it bigger or smaller?

A: Move w in the reverse direction from the slope of the function



Roadmap

- Derivatives

- Basic Concepts of Derivatives
- How to calculate derivatives?
- Partial Derivatives and Gradients
- **Application: the training of machine learning**

- Antiderivatives: Integrals

- Basic Concepts of integrals.
- How to calculate integrals?

Generative vs. Discriminative Classifiers

Suppose we're distinguishing cat from dog images



imagenet



imagenet

Generative Classifier:

- Build a model of what's in a cat image
 - Knows about whiskers, ears, eyes
 - Assigns a *probability* to any image:
 - how cat-y is this image?



Also build a model for dog images

Now given a new image:

Run both models and see which one fits better

Discriminative Classifier

Just try to *distinguish* dogs from cats



Oh look, dogs have collars!
Let's ignore everything else

Components of Discriminative Classifier

Given input/output pairs $(x^{(i)}, y^{(i)})$:

- A **feature representation** of the input. For each input observation $x^{(i)}$, a vector of features $[x_1, x_2, \dots, x_n]$. Feature j for input $x^{(i)}$ is x_j , more completely $x_j^{(i)}$, or sometimes $f_j(x)$.
- A **classification function** that computes \hat{y} , the estimated class, via $p(y|x)$, like the *sigmoid* or *softmax* functions.
- An objective function for learning, like **cross-entropy loss**.
- An algorithm for optimizing the objective function: **gradient descent**.

Example of Classification Features

- For feature x_i , weight θ_i tells is how important is x_i
 - x_i = "review contains '*awesome*'": $\theta_i = +10$
 - x_j = "review contains '*abysmal*'": $\theta_j = -10$
 - x_k = "review contains '*mediocre*'": $\theta_k = -2$

These weights are just assigned as examples. The real assignments may vary in the real training.

*In the task to predict
the positive sentiment*

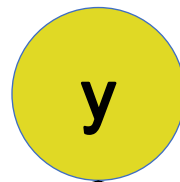


Linear Regression

$$y = h_{\theta}(x) = \sigma(\theta^T x)$$

where $\sigma(a) = a$

Output



θ : the weights
over features

x : the features

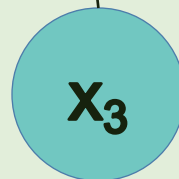
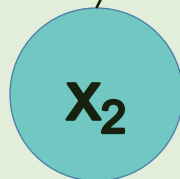
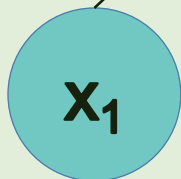
θ_1

θ_2

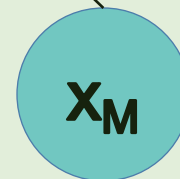
θ_3

θ_M

Input



...



A probabilistic classifier?

$z = \theta^T x$ is a number, and we want to use a function of z that goes from 0 to 1

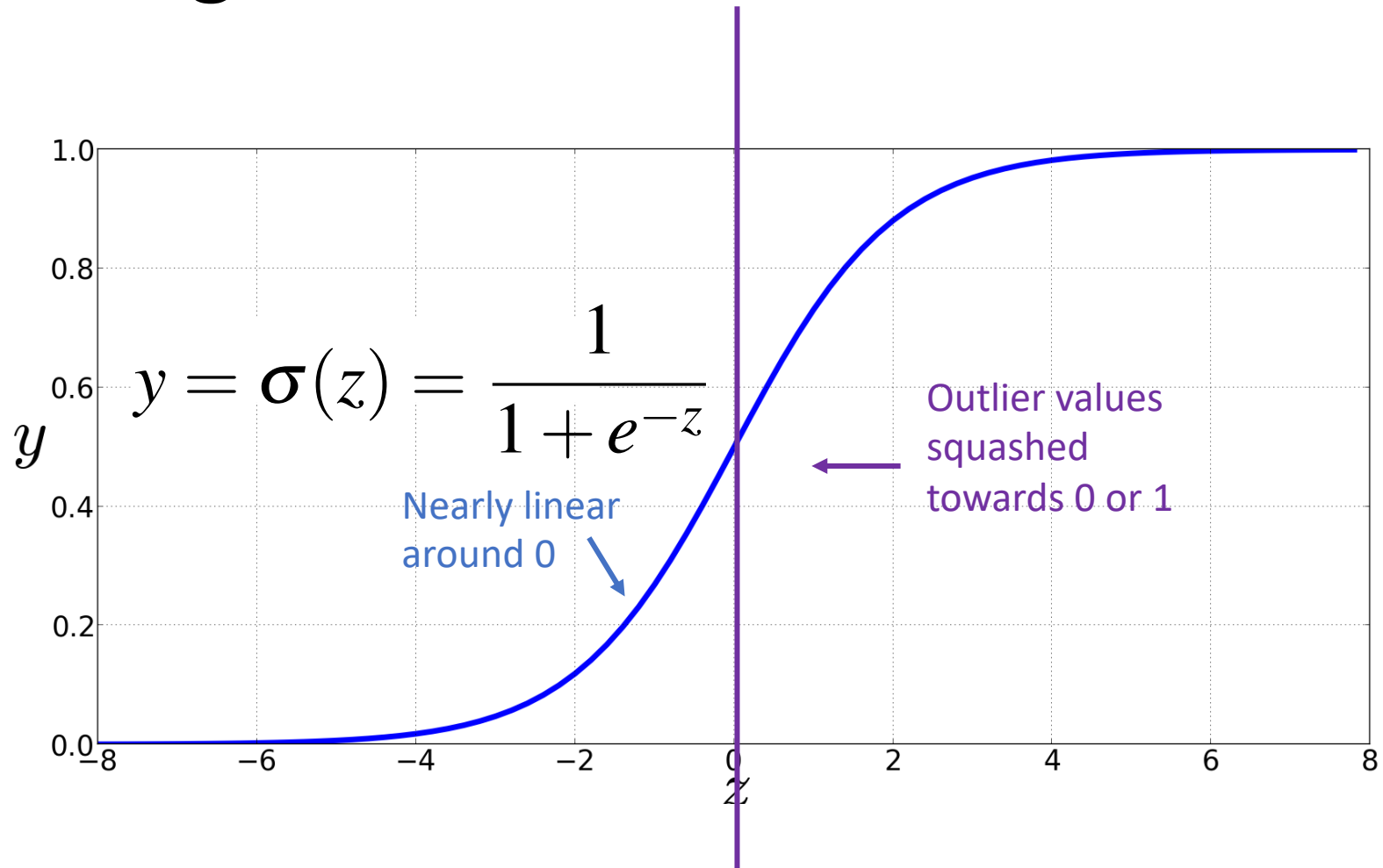
- We need to formalize “sum is high”.
- We’d like a principled classifier that gives us a probability, just like Naive Bayes did
- We want a model that can tell us:

$$p(y = 1|x; \theta)$$

$$p(y = 0|x; \theta)$$

$$y = \sigma(z) = \frac{1}{1 + e^{-z}}$$

The very useful sigmoid or logistic function

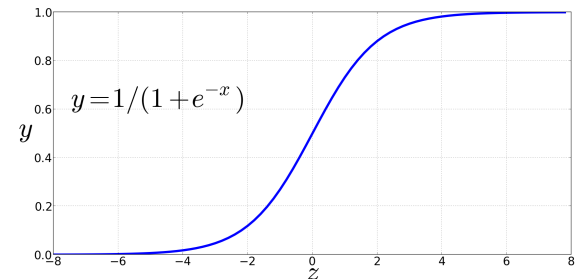
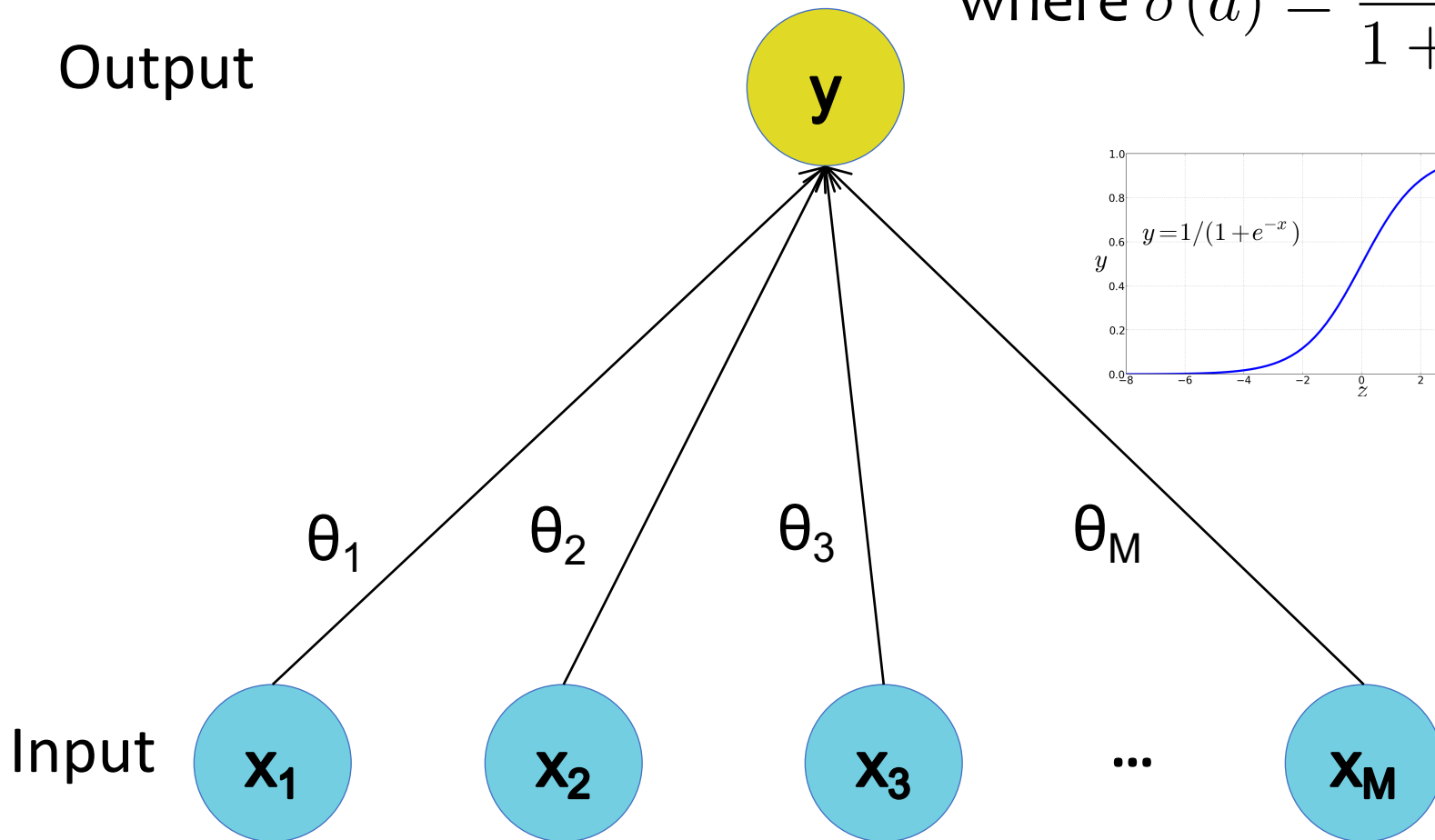


Logistic Regression

$$y = h_{\theta}(x) = \sigma(\theta^T x)$$

$$\text{where } \sigma(a) = \frac{1}{1 + \exp(-a)}$$

Output



A Recipe for Machine Learning

1. Given training data:

$$\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$$

3. Define goal:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^N \ell(f_{\boldsymbol{\theta}}(\mathbf{x}_i), \mathbf{y}_i)$$

2. Choose each of these:

– Decision function

$$\hat{\mathbf{y}} = f_{\boldsymbol{\theta}}(\mathbf{x}_i)$$

– Loss function

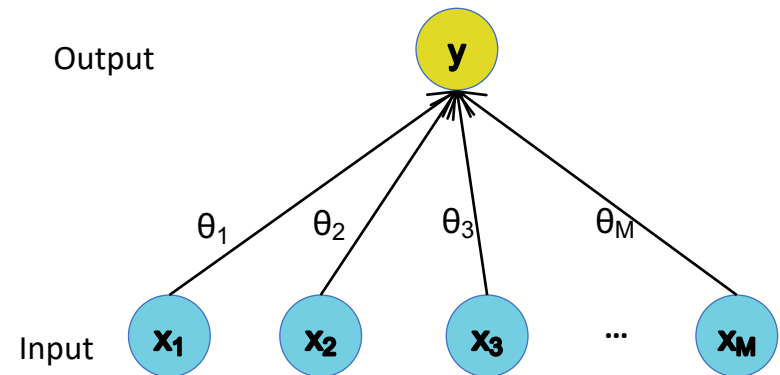
$$\ell(\hat{\mathbf{y}}, \mathbf{y}_i) \in \mathbb{R}$$

4. Train with SGD:

(take small steps opposite the gradient)

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta_t \nabla \ell(f_{\boldsymbol{\theta}}(\mathbf{x}_i), \mathbf{y}_i)$$

Backpropagation for gradient calculation



Forward

$$J = y^* \log y + (1 - y^*) \log(1 - y)$$

$$y = \frac{1}{1 + \exp(-a)}$$

$$a = \sum_{j=0}^D \theta_j x_j$$

Backward

$$\frac{dJ}{dy} = \frac{y^*}{y} + \frac{(1 - y^*)}{y - 1}$$

$$\frac{dJ}{da} = \frac{dJ}{dy} \frac{dy}{da}, \quad \frac{dy}{da} = \frac{\exp(-a)}{(\exp(-a) + 1)^2}$$

$$\frac{dJ}{d\theta_j} = \frac{dJ}{da} \frac{da}{d\theta_j}, \quad \frac{da}{d\theta_j} = x_j$$

$$\frac{dJ}{dx_j} = \frac{dJ}{da} \frac{da}{dx_j}, \quad \frac{da}{dx_j} = \theta_j$$

Roadmap

- Derivatives
 - Basic Concepts of Derivatives
 - How to calculate derivatives?
 - Partial Derivatives and Gradients
 - **Application:** the training of machine learning
- Antiderivatives: Integrals
 - Basic Concepts of integrals.
 - How to calculate integrals?

Roadmap

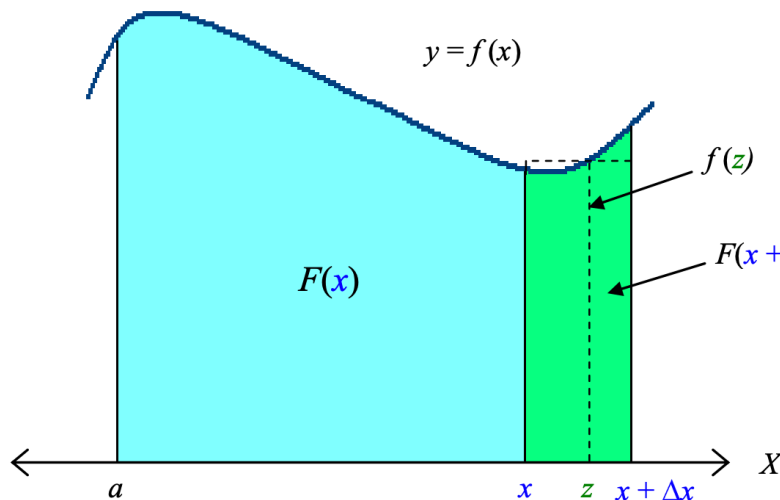
- Derivatives
 - Basic Concepts of Derivatives
 - How to calculate derivatives?
 - Partial Derivatives and Gradients
 - **Application:** the training of machine learning
- Antiderivatives: Integrals
 - Basic Concepts of integrals.
 - How to calculate integrals?

Areas Under Function Graph

- Given $y = f(x)$, *nonnegative* ($f(x) \geq 0$) and *continuous* (with no breaks or jumps).

- Function $F(x) = \int_a^x f(t)dt$ *Antiderivative*

- The area under the graph of $f(x)$ in the interval $[a, x]$



$$F(x + \Delta x) - F(x)$$

= Area under the graph of $f(x)$ in $[x, x + \Delta x]$

$$= f(z) \cdot \Delta x$$

When $\Delta x \rightarrow 0$, $z \rightarrow x$

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x)$$

What are *Integrals*?

- *Definite Integrals*: $F(x) = \int_a^x f(t)dt$
 - Areas under $f(x)$ in the interval of $[a, x]$
 - In this context, $f(x)$ is called the *integrand*.
 - $F(x)$ is an *antiderivative* of $f(x)$.
- *Indefinite integrals*: $\int f(x)dx = F(x) + C$ Arbitrary Constant
 - **Example.** $F(x) = \frac{1}{10}x^{10} + C$ is the general antiderivative of $f(x) = x^9 = F'(x)$
 - $\int x^9 dx = \frac{1}{10}x^{10} + C$

Roadmap

- Derivatives
 - Basic Concepts of Derivatives
 - How to calculate derivatives?
 - Partial Derivatives and Gradients
 - **Application:** the training of machine learning
- Antiderivatives: Integrals
 - Basic Concepts of integrals.
 - How to calculate integrals?

Properties of Integrals

- For any constant c and any integrable function $f(x)$
 - $\int [cf(x)]dx = c \int f(x)dx$
- For any integrable functions $f(x)$ and $g(x)$

- **Sum and Difference Rules:**

$$\int [f(x) \pm g(x)]dx = \int f(x)dx \pm \int g(x) dx$$

- **Power Rule.**
 - **Exponential Rule.**
 - **Chain Rule.**

Properties of Integrals (cont.)

- **Power Rule:**

- $\int u^p du = \begin{cases} \frac{u^{p+1}}{p+1} + C, p \neq -1 \\ \ln|u| + C, p = -1 \end{cases}$

- **Exponential Rule:**

- $\int e^u du = e^u + C$

- **Chain Rule:** $\int f(g(x)) dx = F(g(x))$

- **Example.** $\int \underbrace{(x^5 + 2)^9}_{u^9} \underbrace{5x^4 dx}_{du} = \frac{(x^5 + 2)^{10}}{10} + C$

More Examples of Chain Rules

$$\bullet \int \frac{x^2}{\sqrt{1+x^3}} dx$$

$$u^{-\frac{1}{2}} \quad du$$

$$= \frac{1}{3} \int (1+x^3)^{-\frac{1}{2}} \cdot 3x^2 dx$$

$$= \frac{1}{3} \cdot 2(1+x^3)^{\frac{1}{2}} + C$$

$$\frac{1}{\frac{1}{2}} \frac{u^{\frac{1}{2}}}{\frac{1}{2}} + C$$

$$\bullet \int \frac{x^2}{1+x^3} dx$$

$$u^{-1} \quad du$$

$$= \frac{1}{3} \int (1+x^3)^{-1} \cdot 3x^2 dx$$

$$= \frac{1}{3} \cdot \ln |1+x^3| + C$$

$$\ln |u| + C$$

Fundamental Theorem of Calculus

- How to calculate *definite integrals*:

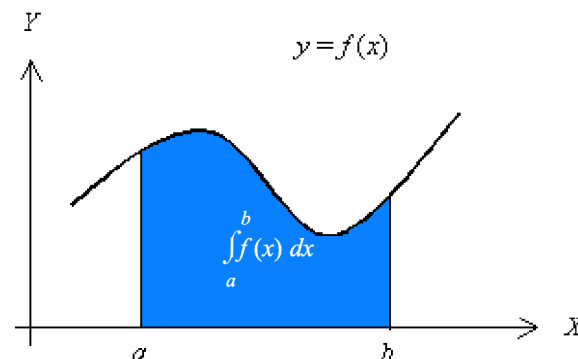
- $\int_a^b f(x) dx = F(b) - F(a)$

- Example.**

- $\int_0^1 x^3 (1 - x^4)^2 dx$

$$= -\frac{1}{4} \int_{x=0}^{x=1} \underbrace{(1 - x^4)^2}_{u^2} \underbrace{(-4)x^3}_{du} dx$$

$$= -\frac{1}{4} \int_{u=1}^{u=0} u^2 du = \frac{1}{4} \int_0^1 u^2 du = \frac{1}{4} \left[\frac{u^3}{3} \right]_0^1 = \frac{1}{12}$$



A slide to take away



Chain Rule:
Super Useful!

- How to calculate derivatives?
 - **Tip1:** Remember the derivatives for the component functions, e.g., $y = x^r$, $y = e^x$, $y = \ln x$, $y = c$, etc.
 - **Tip2:** Remember the properties for *sum*, *product*, *quotient*, and *chain rules*.
 - **Tip3:** Consider the function as the operation results of some component functions, e.g., $y = e^{-\frac{1}{2}x^2}$
- How to use a function to model a problem and how to use gradient descent to find optimal solutions.
- How to calculate integrals?
 - **Tip1:** Consider the antiderivatives first! (Remember the antiderivatives for the important components).
 - **Tip2:** Apply the integral properties for *sum*, *chain rule*, etc.