



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY
BY EVERYONE'S SMART UNIVERSITY

Mathematical Foundations For Data Science and AI

Lecture 11

AI102

Final Revision

By
Dr. Khaled ELMenshawy

www.midocean.ae

Derivatives of Power Functions

$$f(x) = x^{\textcircled{2}}$$

$$f'(x) = 2x^{\textcircled{1}}$$

$$f(x) = x^{\textcircled{3}}$$

$$f'(x) = 3x^{\textcircled{2}}$$

$$f(x) = x^{\textcircled{-1}}$$

$$f'(x) = (-1)x^{\textcircled{-2}}$$

General derivative of power function:

$$f(x) = xn \quad \longrightarrow \quad f'(x) = \frac{d}{dx} f(x) = nx^{n-1}$$

Exercises

What are the derivatives of the following functions?

$$f(x) = 2x^2$$

$$f'(x) = \dots \dots$$

$$f(x) = -x^3$$

$$f'(x) = \dots \dots$$

$$f(x) = 10x^2$$

$$f'(x) = \dots \dots$$

$$f(x) = -x^{-2}$$

$$f'(x) = \dots \dots$$

Summary

Function	Derivatives
$f(x) = \sin(x)$	$f'(x) = \cos(x)$
$f(x) = \cos(x)$	$f'(x) = -\sin(x)$
$f^{-1}(x)$	$\frac{d}{dx} f^{-1}(x) = \frac{1}{f'(x)}$
$f(x) = e^x$	$f'(x) = e^x$
$f(x) = \log(x)$	$\frac{d}{dx} \log(x) = \frac{1}{x}$
$f(x) = cg(x)$	$f'(x) = cg'(x)$
$f(x) = h(x) + g(x)$	$f'(x) = h'(x) + g'(x)$
$f(x) = h(x) \times g(x)$	$f'(x) = g(x)h'(x) + h(x)g'(x)$
$f(g(h(x)))$	$\frac{d}{dx} f(g(h(x))) = \frac{df}{dg} \times \frac{dg}{dh} \times \frac{dh}{dx}$

How to Find the Inverse of a Function?

1. The x and y values will switch in an inverse function, so replace the $f(x)$ notation with y .

$$f(x) = \sqrt{2x - 3} \Rightarrow y = \sqrt{2x - 3}$$

2. Next, replace every x in the equation with a y , and every y with an x .

$$y = \sqrt{2x - 3} \Rightarrow x = \sqrt{2y - 3}$$

3. Solve the new equation for y .

$$x^2 = (\sqrt{2y - 3})^2 \rightarrow x^2 = 2y - 3 \rightarrow 2y = x^2 + 3$$

4. Replace the y with the $f^{-1}(x)$ notation.

$$\therefore y = \frac{x^2 + 3}{2}$$

$$f^{-1}(x) = \frac{x^2 + 3}{2}$$

Examples

Find $f^{-1}(x)$ for the following functions:

$$f(x) = -3x - 7$$

$$y = -3x - 7$$

$$x = -3y - 7$$

$$-3y = x + 7$$

$$y = \frac{x + 7}{-3}$$

$$\therefore f^{-1}(x) = \frac{x + 7}{-3}$$

$$f(x) = \frac{1}{5}x + 8$$

$$y = \frac{1}{5}x + 8$$

$$x = \frac{1}{5}y + 8$$

$$y = 5(x - 8)$$

$$\therefore f^{-1}(x) = 5(x - 8)$$

Exercises

Find $f^{-1}(x)$ for the following functions:

$$f(x) = 2x + 1$$

$$f(x) = -\frac{2}{3}x - 5$$

$$f(x) = -\frac{1}{2}(x + 3)$$

Exercise

The following table lists the values of functions g and h , and of their derivatives, g' and h' , for $x = -2$

Let function $f(x)$ defined as $g(x) \times h(x)$

What is the value of $f'(-2)$?

x	$g(x)$	$h(x)$	$g'(x)$	$h'(x)$
-2	2	-1	3	4

Exercise

The following table lists the values of functions f and h and of their derivatives, f' and h' , for $x = -1$

Evaluate $[f(x) \times h(x)]$ at $x = -1$

x	$f(x)$	$h(x)$	$f'(x)$	$h'(x)$
-1	-3	-4	1	3

Exercise

Let f be a function such $f(-1) = 3$, and $f'(-1) = 5$

Let g be function $g(x) = \frac{1}{x}$.

Let h be a function defined as $h(x) = f(x) \times g(x)$, What is the value of $h'(-1)$?

Exercise

Let $f(3) = -1$, and $f'(3) = 0$

Let $g(x) = x^3$.

Evaluate $\frac{d}{dx} [f(x) \times g(x)]$ at $x = 3$

Partial Derivatives

$$f(x, y)$$



Partial derivative notation

$$f_x = \frac{\partial f}{\partial x}$$

*Partial derivative of f
with respect to x*

Partial derivative notation

$$f_y = \frac{\partial f}{\partial y}$$

*Partial derivative of f
with respect to y*

Summary**Solving function of two variables:**

Find partial derivative of f with respect to x

- Treat all other variables as constants. In this case y .
- Differentiate the function using the normal rules of differentiation.

$$f_x = \frac{\partial f}{\partial x} = 2x$$

Find partial derivative of f with respect to y

- Treat all other variables as constants. In this case x .
- Differentiate the function using the normal rules of differentiation.

$$f_y = \frac{\partial f}{\partial y} = 2y$$

Example

$$f(x, y) = 3x^2y^3$$

$$f(x, y) = 3x^2y^3$$

$$f_x = \frac{\partial f}{\partial x} = 3(2x)y^3$$

$$f_x = \frac{\partial f}{\partial x} = 6xy^3$$

$$f(x, y) = 3x^2y^3$$

$$f_x = \frac{\partial f}{\partial x}$$

$$= 3x^2(3y^2)$$

$$f_x = \frac{\partial f}{\partial x} = 9x^2y^2$$

Gradient Descent

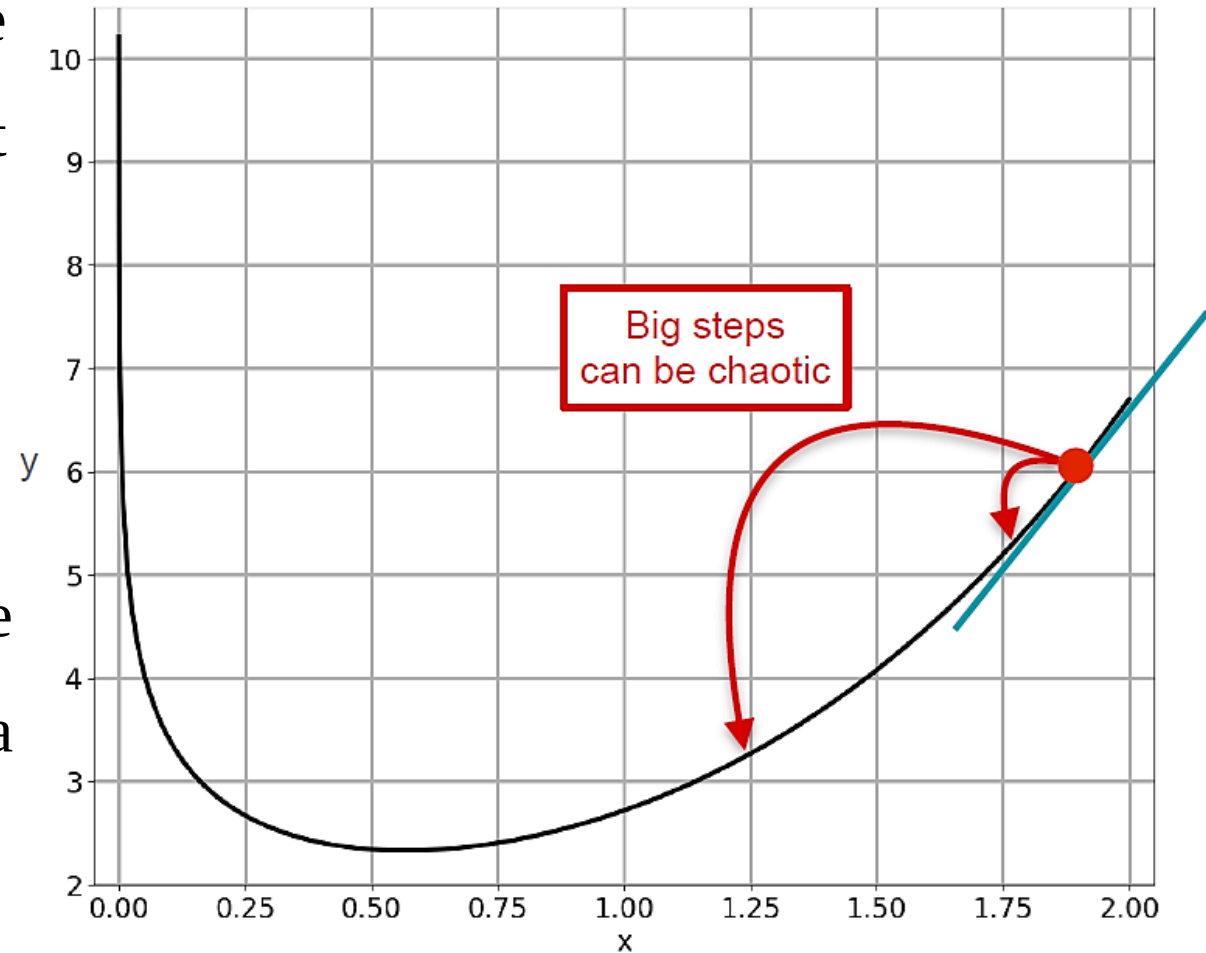
If we want a new point that is closer to the minimum, then that should be the old point minus the slope.

$$\text{New Point} = \text{Old Point} - \text{Slope}$$

$$x_{\text{new}} = x_{\text{old}} - f'(x_{\text{old}})$$

To take small steps so that we do not miss the minimum point, simply multiply the slope by a small number called the “*Learning Rate*”.

$$x_{\text{new}} = x_{\text{old}} - \alpha f'(x_{\text{old}})$$



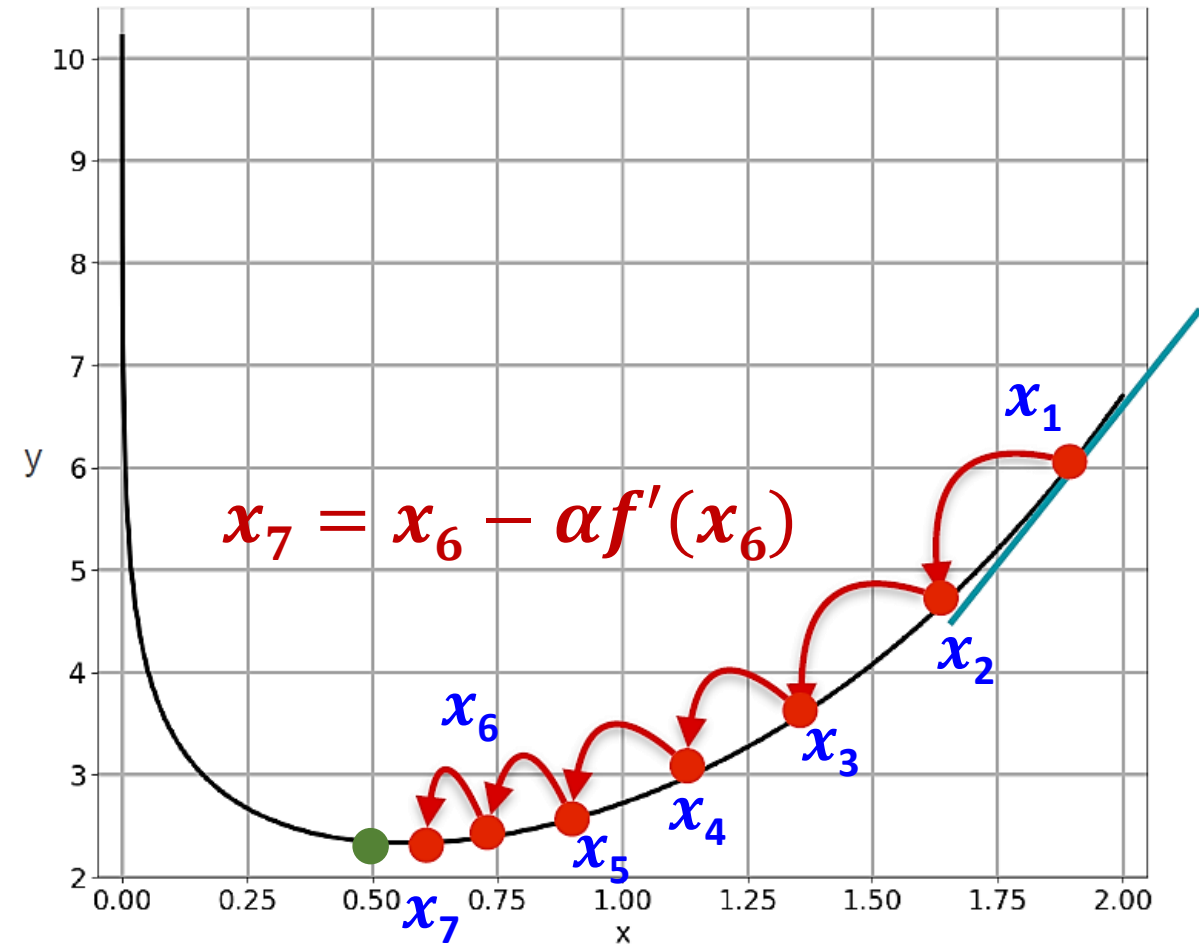
Gradient Descent:

Target: Finding the minimum of $f(x)$

Step (1): Define the learning rate (α) and choose a starting point (x_0)

Step (2): Update $x_n = x_{n-1} - \alpha f'(x_{n-1})$

Step (3): Repeat (Step 2) until the minimum point is reached



Gradient Descent:

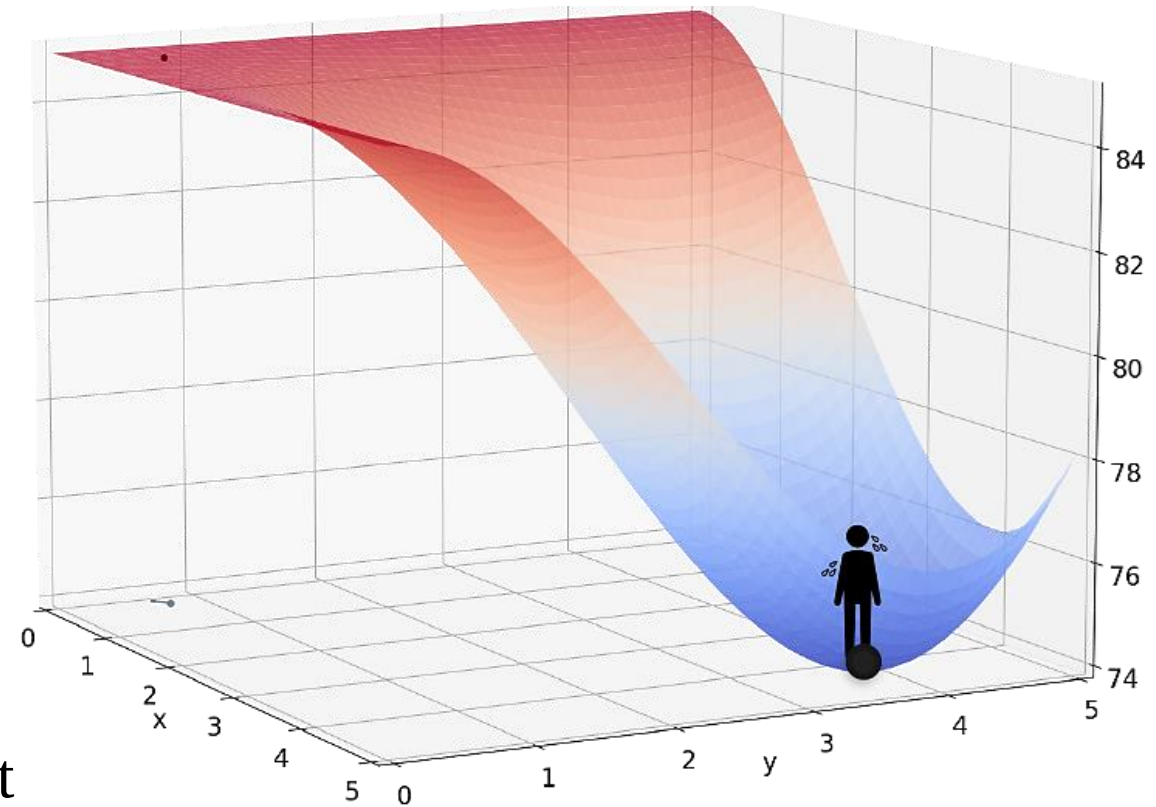
Target: Finding the minimum of $f(x)$

Step (1): Define the learning rate (α) and choose a starting point (x_0, y_0)

Step (2): Update

$$\begin{bmatrix} x_{new} \\ y_{new} \end{bmatrix} = \begin{bmatrix} x_{old} \\ y_{old} \end{bmatrix} - \alpha \nabla f(x_{old}, y_{old})$$

Step (3): Repeat (Step 2) until the minimum point (x, y) is reached

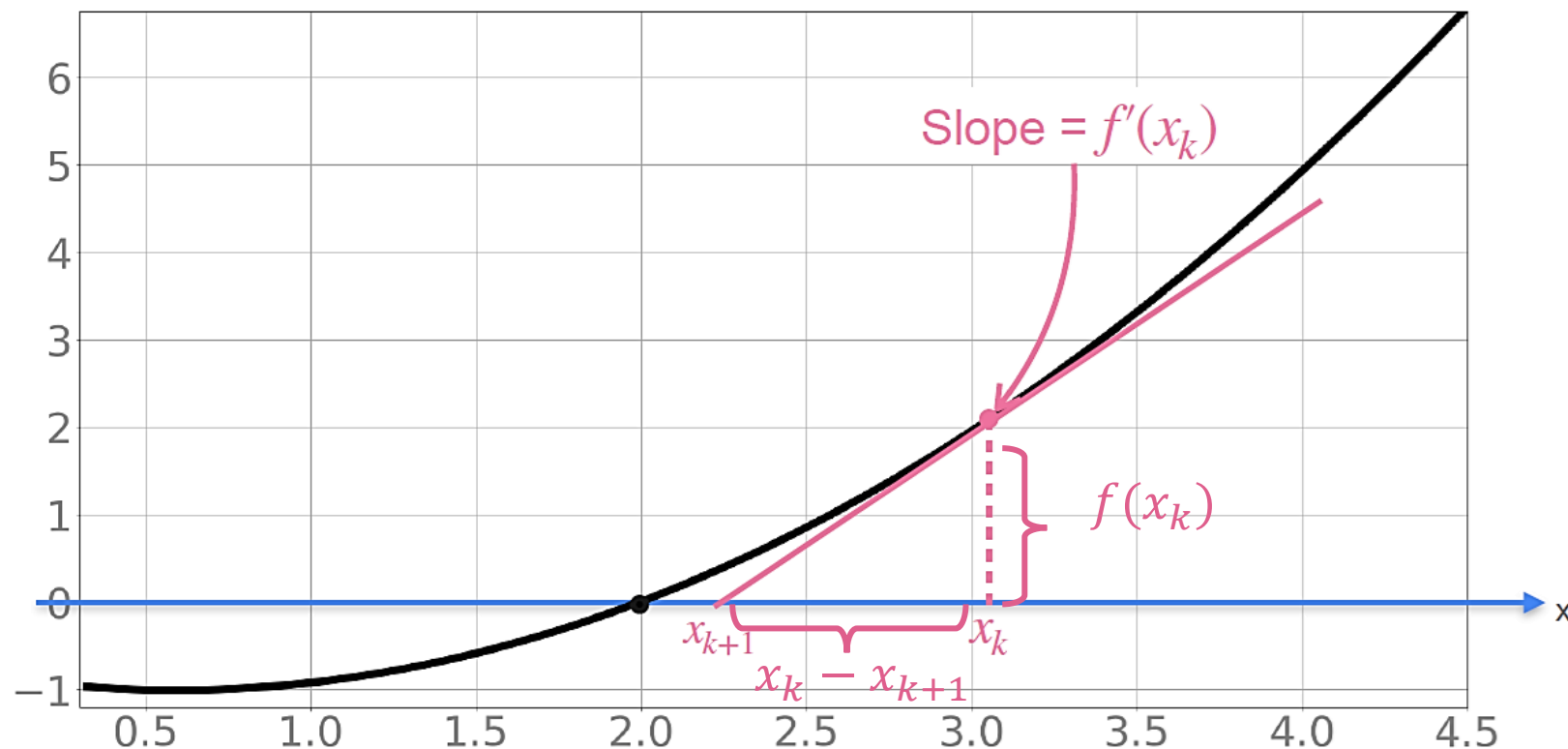


Method	Rule
Square Loss function	$x = \frac{a_1 + a_2 + a_3 + \dots + a_n}{n}$
The Gradient of $f(x, y)$	$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$
Gradient Descent	$x_{new} = x_{old} - \alpha f'(x_{old})$

Rule	Logarithm	Equal to
<i>Product Rule</i>	$\log xy$	$\log x + \log y$
<i>Division Rule</i>	$\log \frac{x}{y}$	$\log x - \log y$
<i>Exponential Rule</i>	$\log x^c$	$c \log x$

Method	Rule
The Gradient of $f(x, y)$	$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$
Gradient Descent	$x_{new} = x_{old} - \alpha f'(x_{old})$

Newton's Method



$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

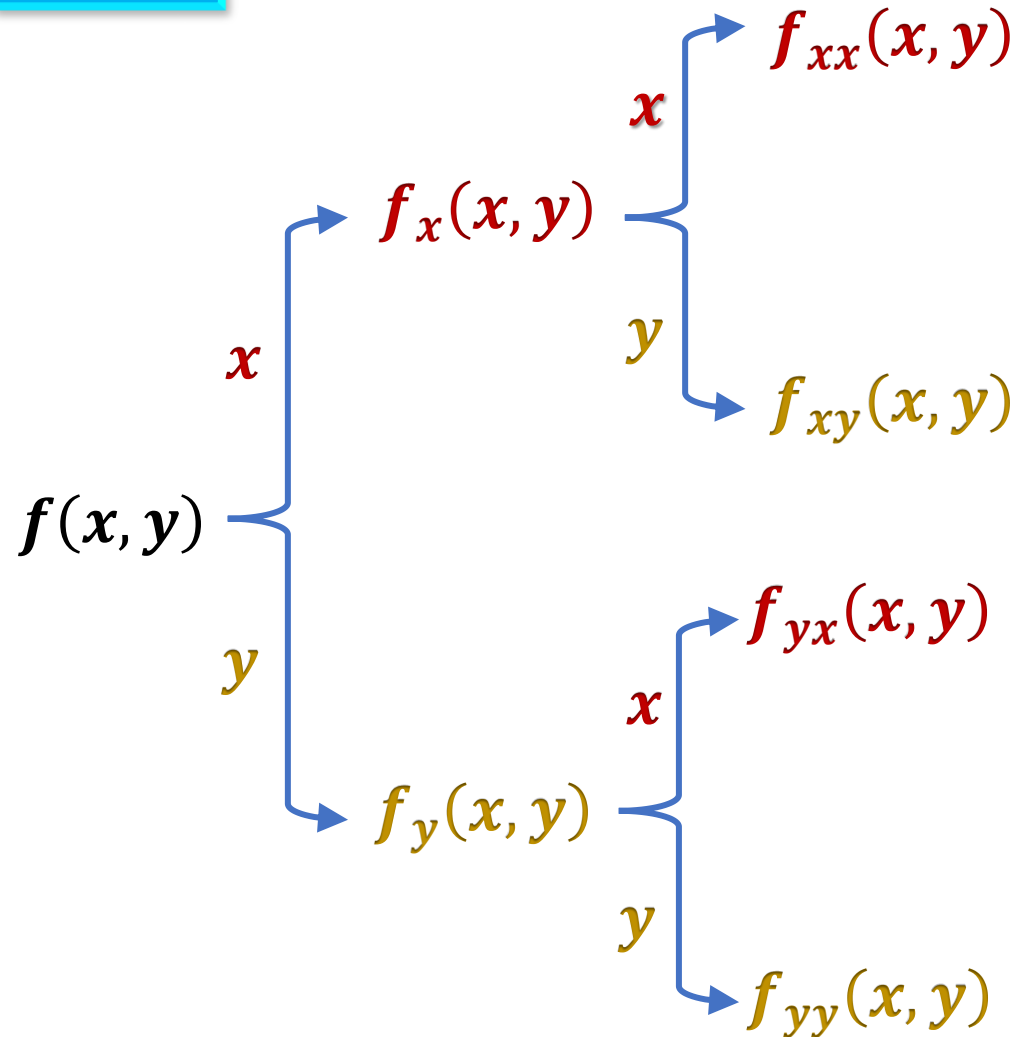
The Hessian

$$f(x, y) = 2x^2 + 3y^2 - xy$$

$f_x(x, y)$ (red) $\rightarrow$ $f_{xx}(x, y)$ (red) $\rightarrow$ 4
 $f_{xy}(x, y)$ (yellow) $\rightarrow$ -1
 $f_y(x, y)$ (yellow) $\rightarrow$ $f_{yx}(x, y)$ (red) $\rightarrow$ -1
 $f_{yy}(x, y)$ (yellow) $\rightarrow$ 6

	Leibniz's notation	Lagrange's notation
Rate of change of $f'_x(x, y)$ with respect to x	$\frac{\partial^2 f}{\partial x^2}$	$f_{xx}(x, y)$
Rate of change of $f'_y(x, y)$ with respect to y	$\frac{\partial^2 f}{\partial y^2}$	$f_{yy}(x, y)$
Rate of change of $f'_x(x, y)$ with respect to y	$\frac{\partial^2 f}{\partial x \partial y}$	$f_{xy}(x, y)$
Rate of change of $f'_y(x, y)$ with respect to x	$\frac{\partial^2 f}{\partial y \partial x}$	$f_{yx}(x, y)$

The Hessian



$$\nabla f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix} \quad \nabla f_y = \begin{bmatrix} f_{yx}(x, y) \\ f_{yy}(x, y) \end{bmatrix}$$

$$H = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

Hessian Matrix

The Hessian

$f(x, y) = 2x^2 + 3y^2 - xy$

First-order partial derivatives:

- $f_x(x, y) = 4x - y$
- $f_y(x, y) = 6y - x$

Second-order partial derivatives:

- $f_{xx}(x, y) = 4$
- $f_{xy}(x, y) = -1$
- $f_{yx}(x, y) = -1$
- $f_{yy}(x, y) = 6$

$$H = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

Hessian Matrix

	One Variable	Two Variable
Function	$f(x)$	$f(x, y)$
First Derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t. x $f_y(x, y)$ Rate of change w.r.t. y $\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$
Second Derivative	$f''(x)$ Rate of change of the rate of change $f(x)$	$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$

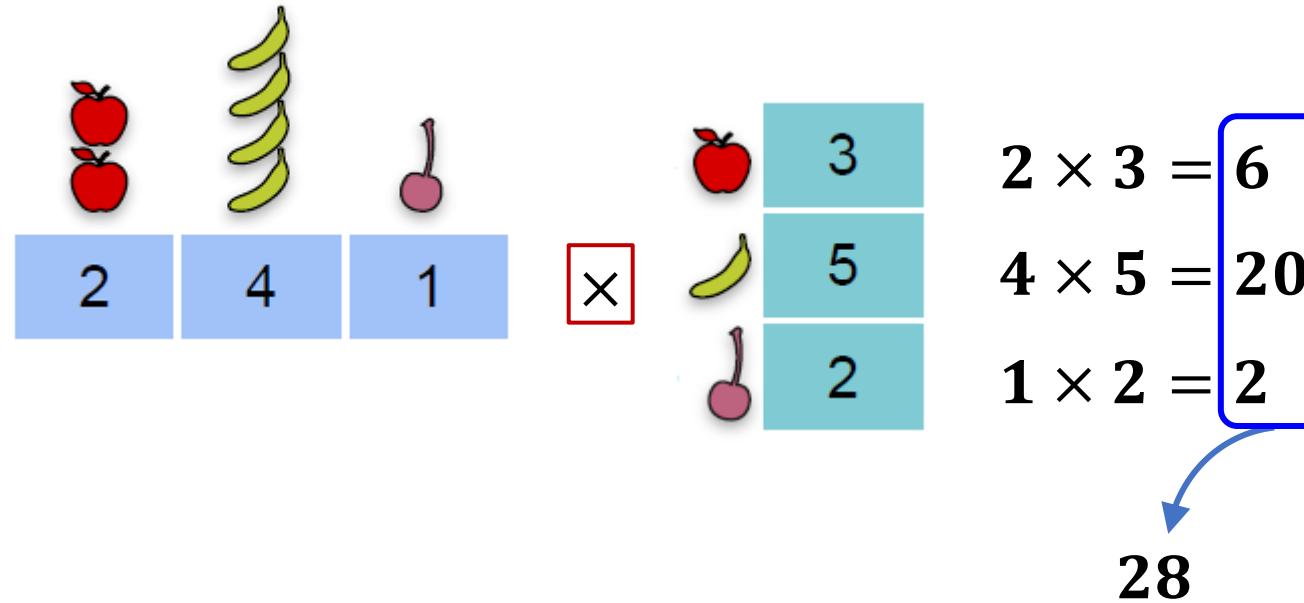
***The Matrix** is a collection of numbers ordered in rows and columns. We would normally denote a matrix with a capital letter, for instance, A*

$$A_{3 \times 3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

} 3 Rows

} 3 Columns

The Dot Product



The diagram illustrates the dot product of two vectors using fruit counts. On the left, a vector is represented by three blue boxes containing the numbers 2, 4, and 1, with two apples above the first box, four bananas above the second, and one cherry above the third. This is followed by a red box containing a multiplication symbol (×). To the right, another vector is shown with three light blue boxes containing the numbers 3, 5, and 2, with one apple to the left of the first box, one banana to the left of the second, and one cherry to the left of the third. To the right of this second vector, the following calculations are shown:

$$\begin{aligned} 2 \times 3 &= 6 \\ 4 \times 5 &= 20 \\ 1 \times 2 &= 2 \end{aligned}$$

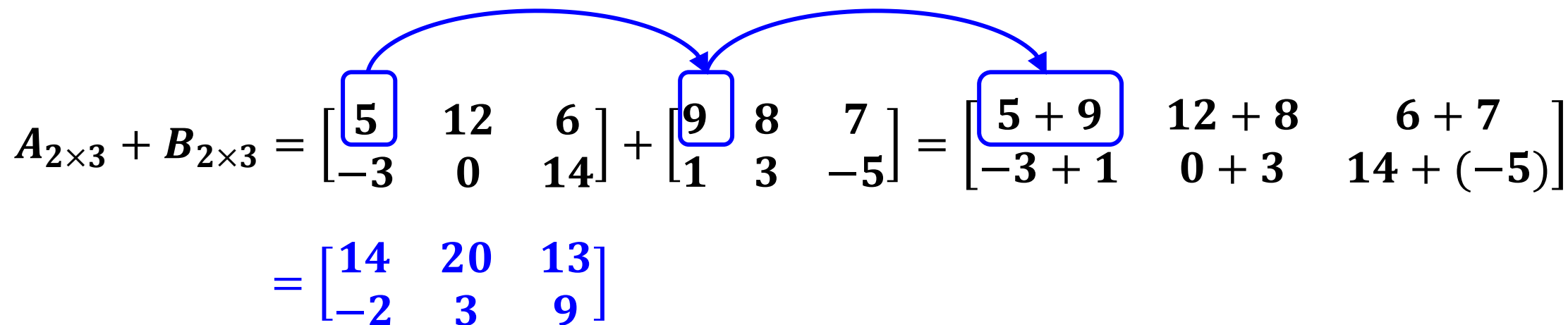
The results 6, 20, and 2 are enclosed in a blue box. A blue arrow points from this box to the number 28, which represents the sum of the products (6 + 20 + 2).

$$\begin{bmatrix} 2 & 4 & 1 \end{bmatrix} \times \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = 28$$

Addition and Subtraction of Matrices

The two matrices must have the same dimensions.

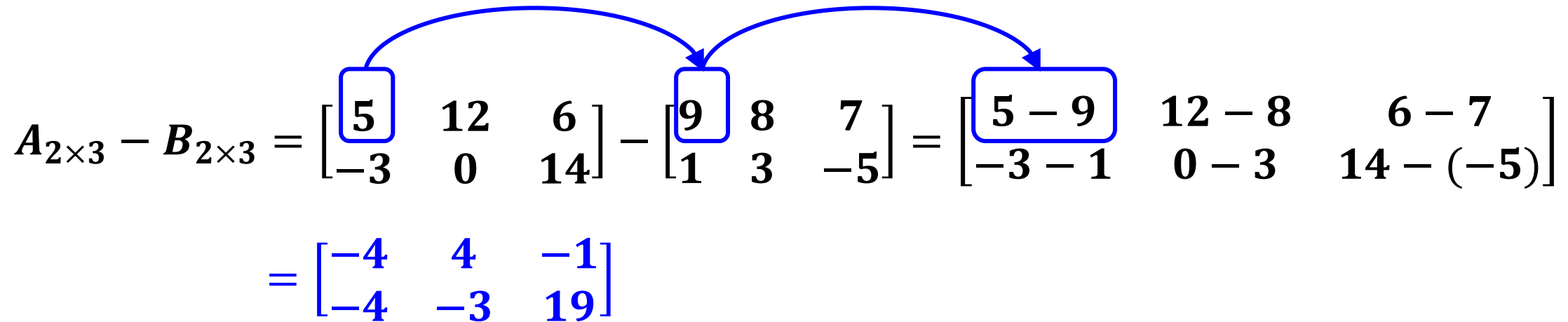
$$A_{2 \times 3} = \begin{bmatrix} 5 & 12 & 6 \\ -3 & 0 & 14 \end{bmatrix} \quad B_{2 \times 3} = \begin{bmatrix} 9 & 8 & 7 \\ 1 & 3 & -5 \end{bmatrix}$$


$$A_{2 \times 3} + B_{2 \times 3} = \begin{bmatrix} 5 & 12 & 6 \\ -3 & 0 & 14 \end{bmatrix} + \begin{bmatrix} 9 & 8 & 7 \\ 1 & 3 & -5 \end{bmatrix} = \begin{bmatrix} 5 + 9 & 12 + 8 & 6 + 7 \\ -3 + 1 & 0 + 3 & 14 + (-5) \end{bmatrix}$$
$$= \begin{bmatrix} 14 & 20 & 13 \\ -2 & 3 & 9 \end{bmatrix}$$

Addition and Subtraction of Matrices

The two matrices must have the same dimensions.

$$A_{2 \times 3} = \begin{bmatrix} 5 & 12 & 6 \\ -3 & 0 & 14 \end{bmatrix} \quad B_{2 \times 3} = \begin{bmatrix} 9 & 8 & 7 \\ 1 & 3 & -5 \end{bmatrix}$$


$$A_{2 \times 3} - B_{2 \times 3} = \begin{bmatrix} 5 & 12 & 6 \\ -3 & 0 & 14 \end{bmatrix} - \begin{bmatrix} 9 & 8 & 7 \\ 1 & 3 & -5 \end{bmatrix} = \begin{bmatrix} 5 - 9 & 12 - 8 & 6 - 7 \\ -3 - 1 & 0 - 3 & 14 - (-5) \end{bmatrix}$$
$$= \begin{bmatrix} -4 & 4 & -1 \\ -4 & -3 & 19 \end{bmatrix}$$

Using the same concept, the dot product, we can multiply two matrices. We can only multiply a matrix of dimensions $m \times n$ with a matrix of dimensions $n \times k$.

The second dimension of the first matrix must match the first dimension of the second matrix.

1. *Matrices are nothing more than a collection of vectors.*
2. *When we have a dot product, multiply a row vector times a column vector.*

$$\begin{bmatrix} 5 & 12 & 6 \\ -3 & 0 & 14 \end{bmatrix} \times \begin{bmatrix} 2 & -1 \\ 8 & 0 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 124 & -5 \\ 36 & 3 \end{bmatrix}$$

***The transpose of a matrix** is found by interchanging its rows into columns or columns into rows. The transpose of the matrix is denoted by using the letter “T” in the superscript of the given matrix.*

$$A = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \longrightarrow A^T = [3 \quad 4 \quad 5]^T$$

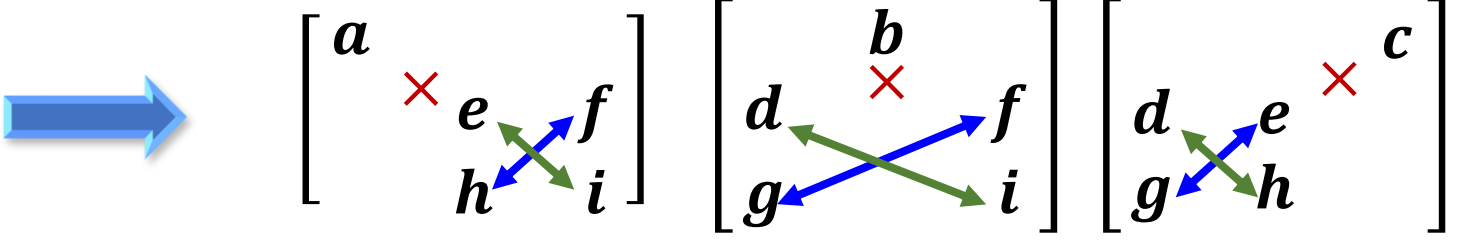
$$A_{2 \times 3} = \begin{bmatrix} 5 & 12 & 6 \\ -3 & 1 & 14 \end{bmatrix} \longrightarrow A_{3 \times 2}^T = \begin{bmatrix} 5 & -3 \\ 12 & 1 \\ 6 & 14 \end{bmatrix}^T$$

*The **identity matrix** is the matrix that when multiplied by any other matrix it gives the same matrix.*

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

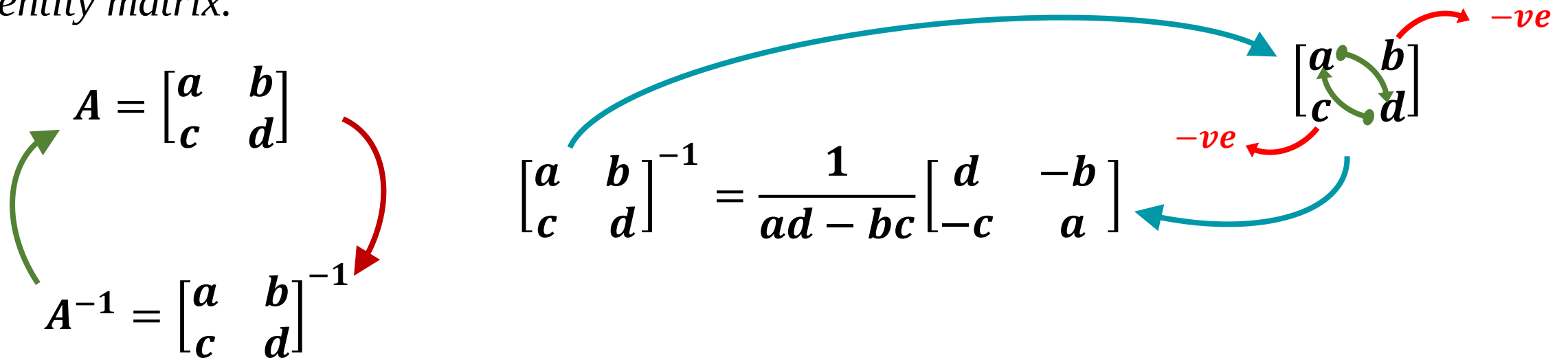
The determinant is a special number that can be calculated from a matrix. The matrix has to be square (same number of rows and columns).

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow |A| = ad - bc$$


$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \Rightarrow \begin{bmatrix} a & \times e & f \\ & h & i \end{bmatrix} \begin{bmatrix} & b & \times \\ d & & f \\ g & & i \end{bmatrix} \begin{bmatrix} & & c & \times \\ d & e & \\ g & h & \end{bmatrix}$$


$$\Rightarrow |A| = a(ei - fh) - b(di - fg) + c(dh - eg)$$

Inverse of a Matrix is a matrix that on multiplying with the original matrix result in the identity matrix.



$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

In other words: swap the positions of **a** and **d**, put negatives in front of **b** and **c**, and divide everything by **ad - bc** (the determinant of the matrix)

$$\begin{aligned}\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}^{-1} &= \frac{1}{4 \times 6 - 7 \times 2} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}\end{aligned}$$

How do we know this is the right answer?

$$\begin{aligned}\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix} \times \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix} \\ &= \begin{bmatrix} 4 \times 0.6 + 7 \times -0.2 & 4 \times -0.7 + 7 \times 0.4 \\ 2 \times 0.6 + 6 \times -0.2 & 2 \times -0.7 + 6 \times 0.4 \end{bmatrix} \\ &= \begin{bmatrix} 2.4 - 1.4 & -2.8 + 2.8 \\ 1.2 - 1.2 & -1.4 + 2.4 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\end{aligned}$$

Non-zero determinants mean that the matrix has an inverse and a zero determinant, that the matrix does not have an inverse.

Subject	Description
<i>The complement rule</i>	$P(A') = 1 - P(A)$
<i>The Sum of probability (Disjoint-Events)</i>	<i>The sum of all probabilities</i>
<i>Sum of Probabilities (Joint-Events)</i>	<i>The sum of probabilities subtracted from Joint events</i>
<i>Conditional Probability</i>	$P(A \cap B) = P(A) \times P(B A)$
<i>Bayes Theorem</i>	$P(A B) = \frac{P(A) \times P(B A)}{P(A) \times P(B A) + P(A') \times P(B A')}$
<i>Bernoulli Distribution</i>	$X \sim \text{Bernoulli "Bern"} (p)$
<i>Binomial Distribution</i>	$X \sim \text{Binomial "B"} (n, p)$
<i>Poisson Distribution</i>	$P(y) = \frac{\lambda^y e^{-\lambda}}{y!}$

Subject	Description
<i>The Mean</i>	$\frac{1}{n} \sum_{i=1}^n X_i$
<i>Variance σ^2</i>	<i>The average of the squared differences from the Mean</i>
<i>The Standard Deviation σ</i>	$\sqrt{\sigma^2}$
<i>Normal distribution</i>	$X \sim \text{Normal "N"} (\mu, \sigma^2)$
<i>Standardization</i>	$\frac{X - \mu}{\sigma}$
<i>Z-Score</i>	$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

Subject	Description
<i>T-distribution</i>	$T = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$
<i>Estimate of SD</i>	$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$

