Day - 87, Feb 25, 2025 (Magh 13, 2081)

1. The Second Derivative (Newton's Method and Leibniz Notation)
2. Hessian and Hessian Matrix
3. Hessian Matrix and Concavity
4. Newton's Method for two variables
5. Graphs and Conceptual figures, Matrices and Eigen Values
6. Examples on the Newton's Method [Optimization Problems]

Source: Calculus for Machine Learning and Data Science offered by DeepLearning AI in Coursera

[Derivative of a derivative - Second Derivative]

Newton's method: $x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$?? → ??

Second Derivative

Leibniz Notation: $\frac{d^2 f(x)}{dx^2} = \frac{d}{dx}\left(\frac{d f(x)}{dx}\right)$

Lagrange Notation: $f''(x)$ as Second Derivative and $f'(x)$ is the first Derivative.

$x \to$ distance

$v \to$ velocity

$a \to$ acceleration

$$\frac{dx}{dt}$$

$$\frac{dv}{dt} = \frac{d^2x}{dt^2}$$

$$v \to \frac{dx}{dt}$$

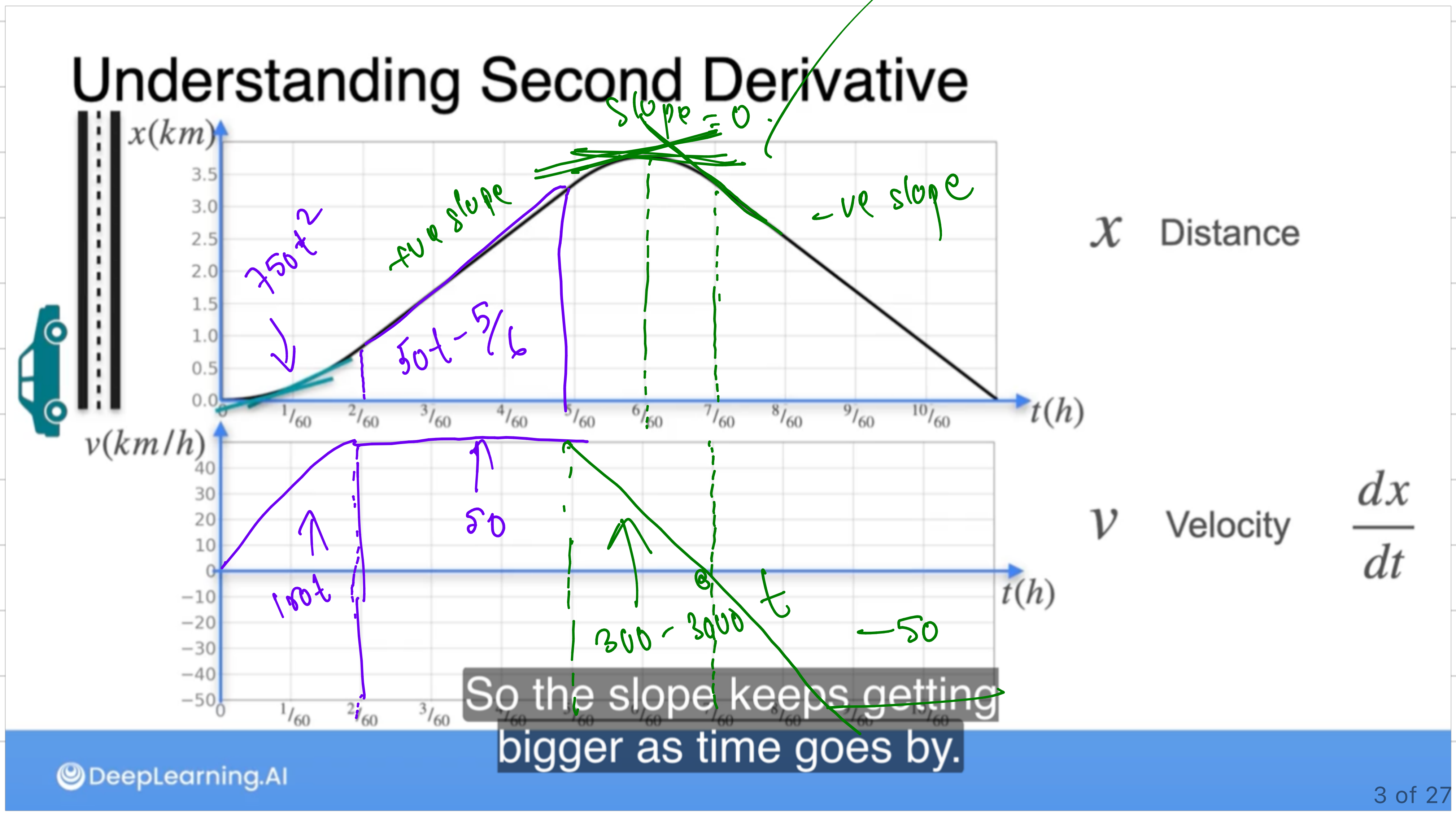
$$a = \frac{dv}{dt}$$

$$-1500t^2 + 30t - 11.16$$

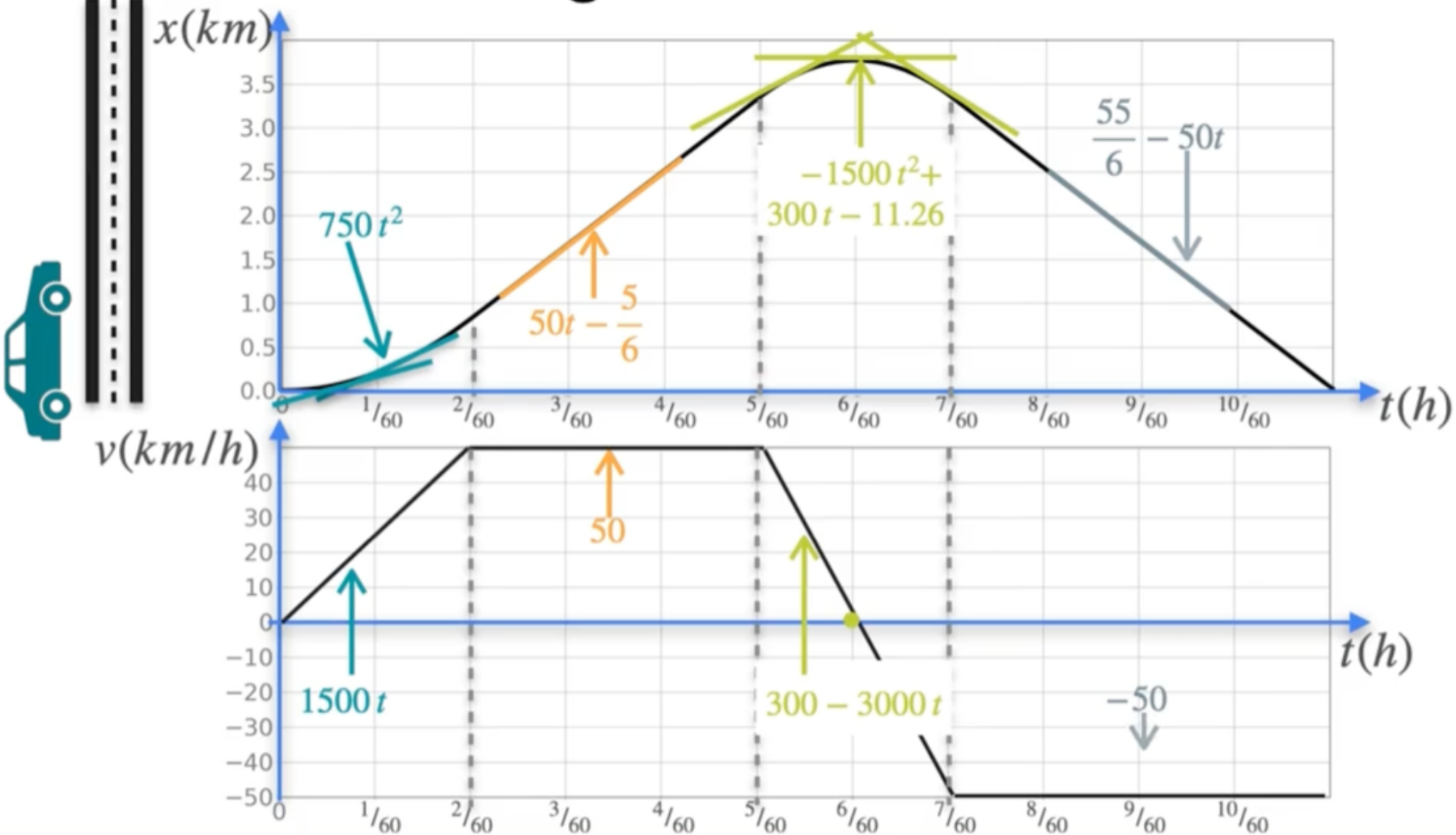
v velocity $\frac{dx}{dt}$

a acceleration $\frac{dv}{dt}$

$t(h)$



Understanding Second Derivative



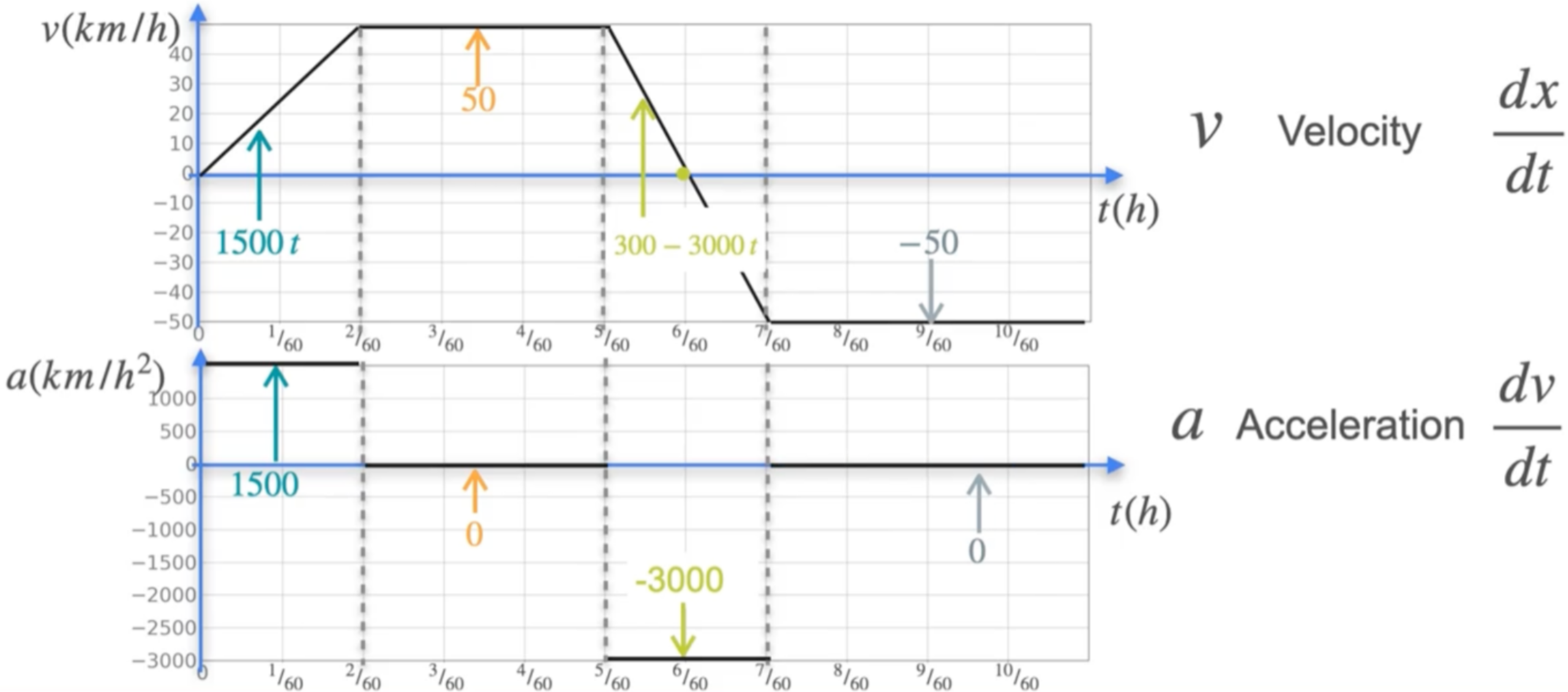
x Distance

v Velocity $\frac{dx}{dt}$

DeepLearning.AI

Q. if the second derivative of a function f(x) i.e. $f''(x)$ or $\frac{d^2}{dx^2} f(x)$ is equal to 0 for every value of x, what can we say about f(x)?

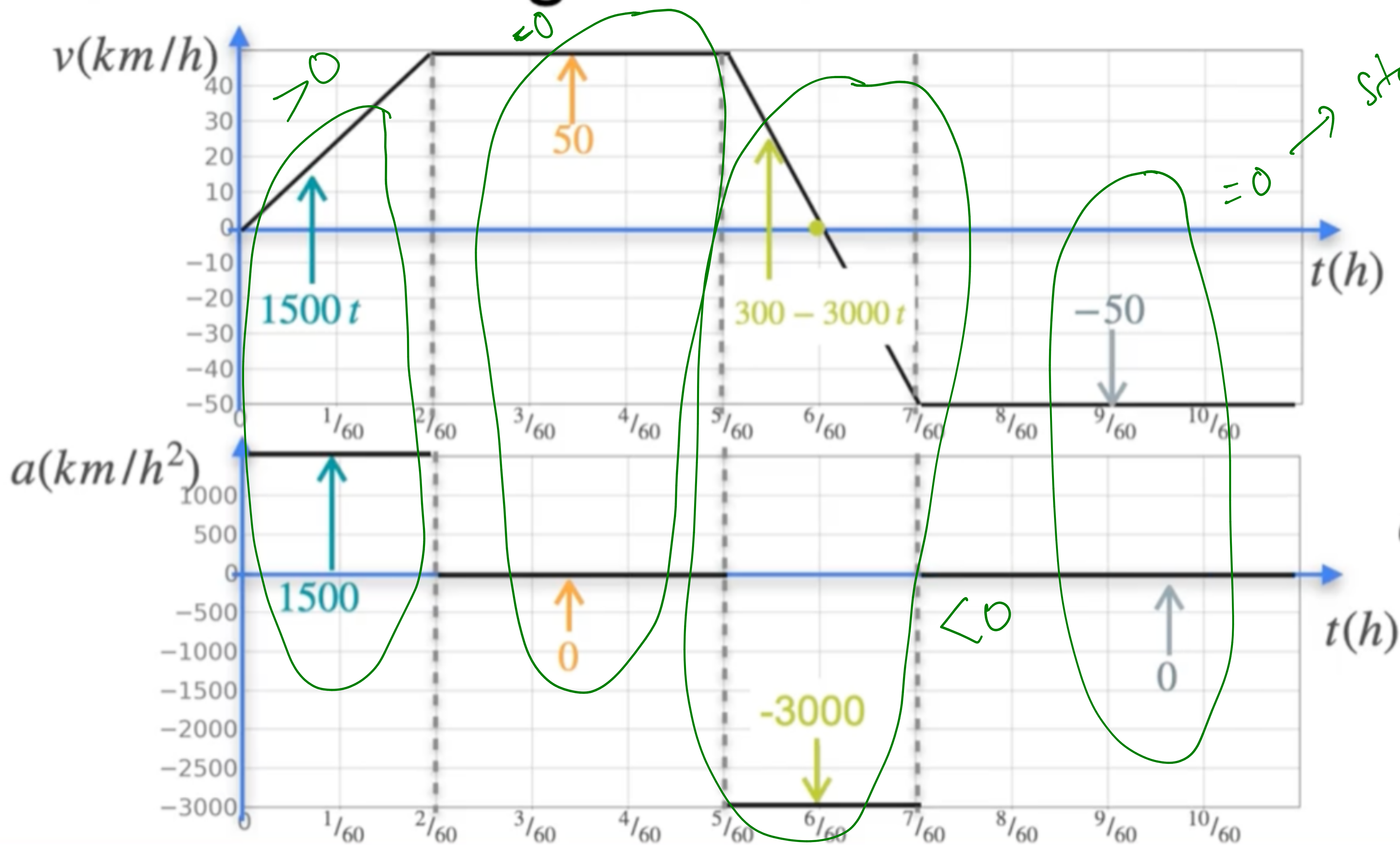
Understanding Second Derivative



DeepLearning.AI

→ We can say f(x) is a line because f'(x) is constant means it must be a line (either horizontal or not)!

Understanding Second Derivative

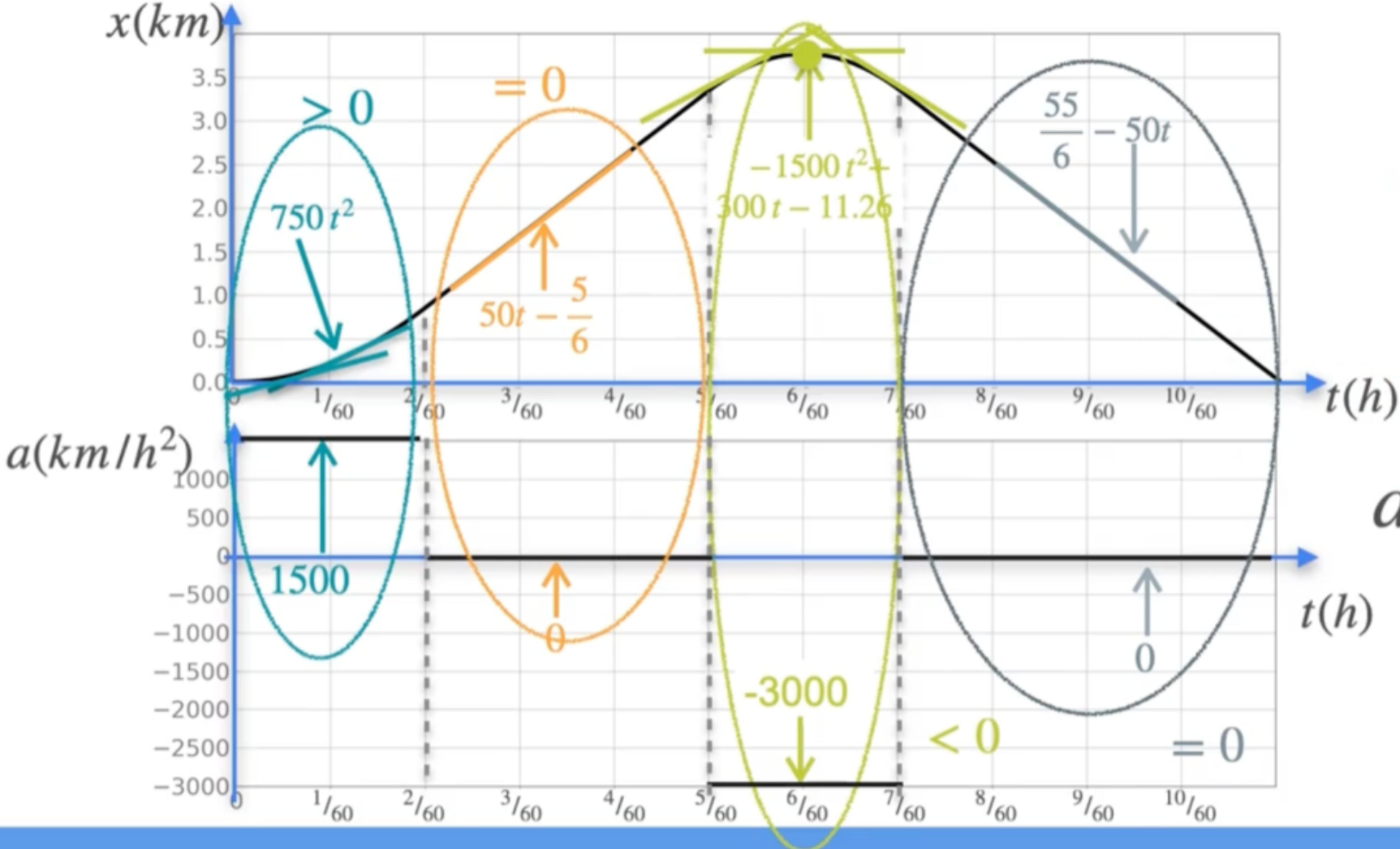


v Velocity $\frac{dx}{dt}$

a Acceleration $\frac{dv}{dt}$

DeepLearning.AI

Understanding Second Derivative



x Distance

Second derivative tells us about the curvature

a Acceleration $\frac{d^2x}{dt^2}$

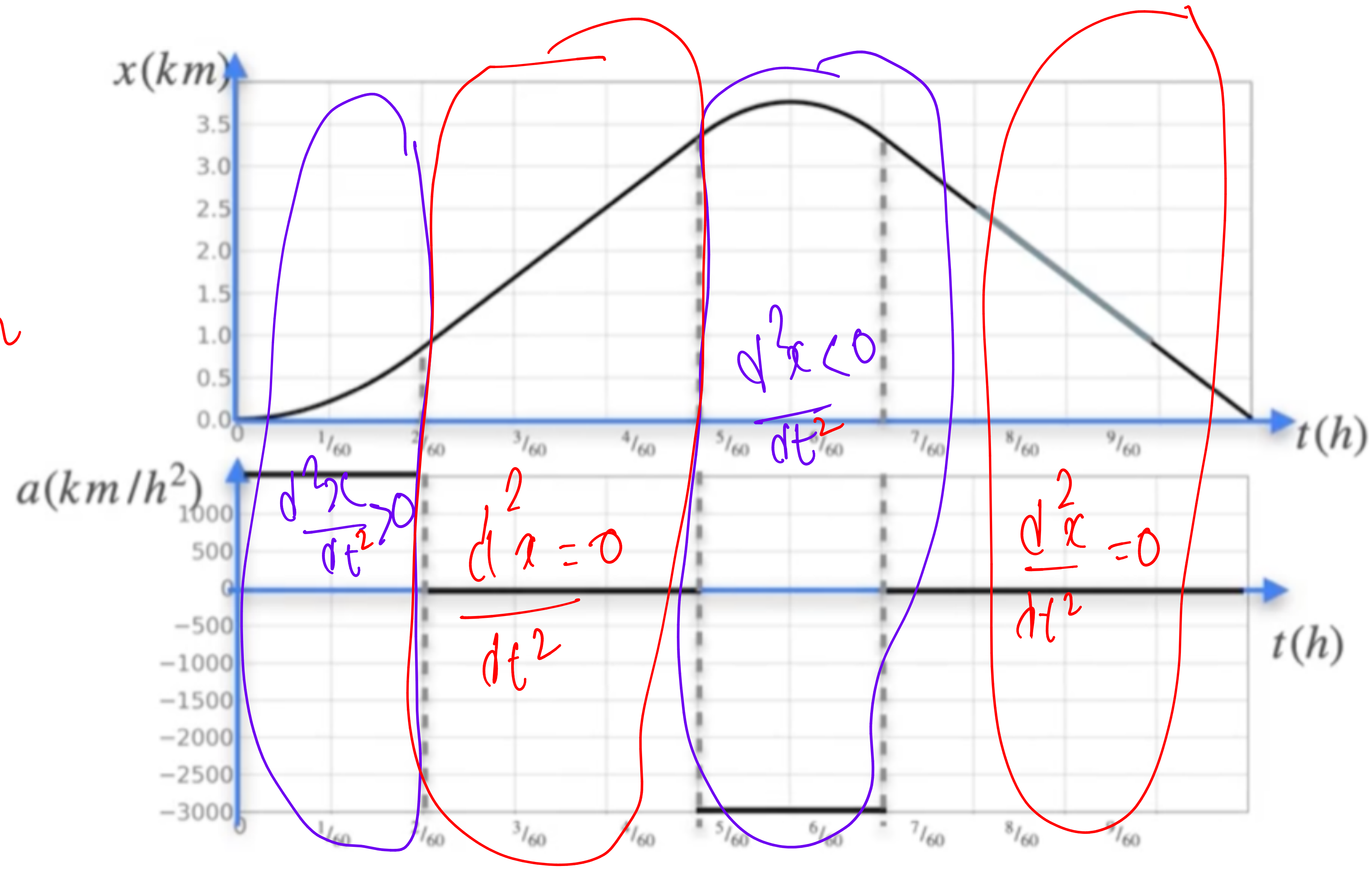
DeepLearning.AI

Curvature

$\frac{d^2x}{dt^2} > 0$ Concave up or Convex

$\frac{d^2x}{dt^2} < 0$ Concave down

$\frac{d^2x}{dt^2} = 0$ Need More Information!



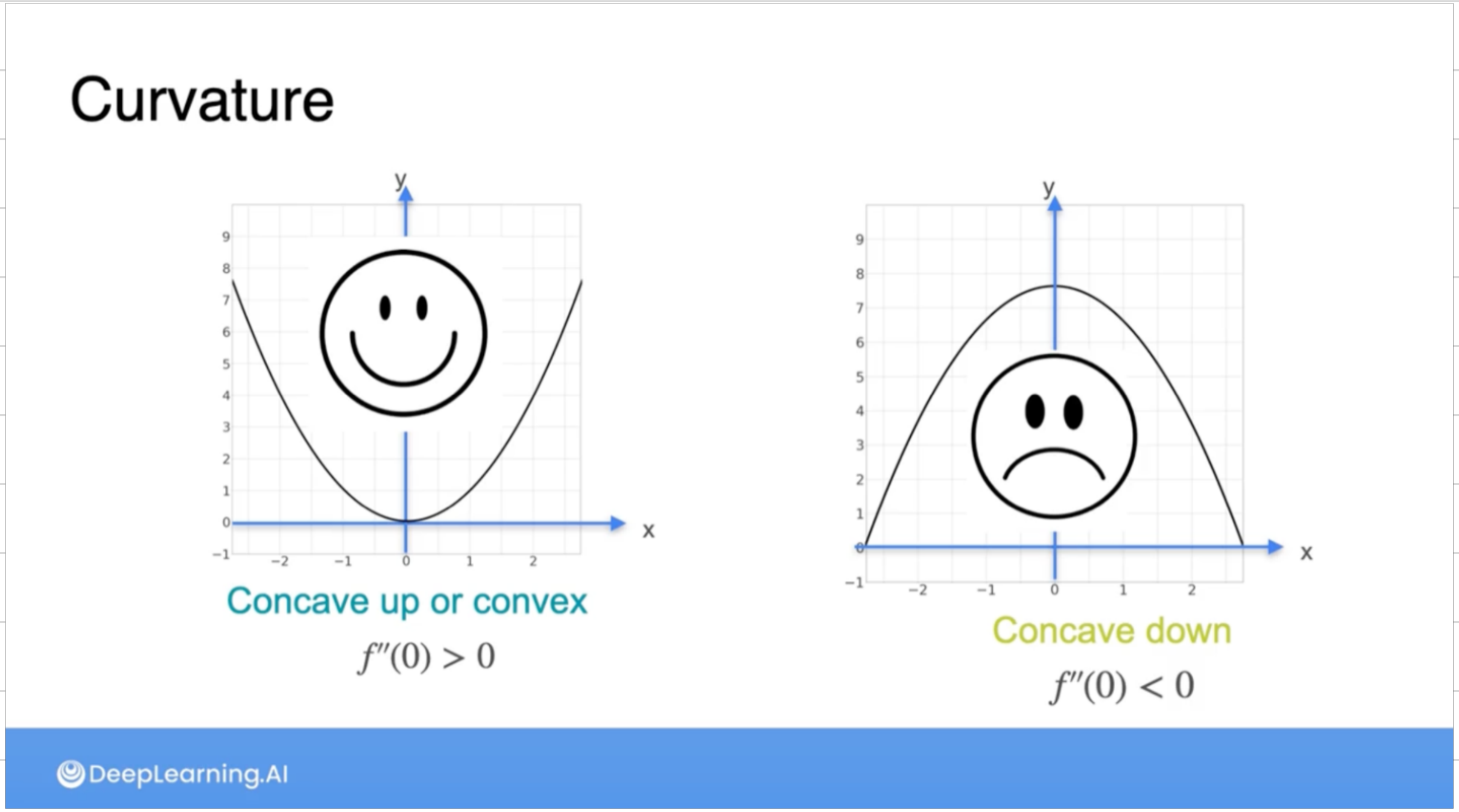
Curvature

$$\frac{d^2x}{dt^2} > 0$$ Concave up or convex

$$\frac{d^2x}{dt^2} < 0$$ Concave down

$$\frac{d^2x}{dt^2} = 0$$ Need more information

DeepLearning.AI



$\frac{d^2 f}{dt^2}$ tells max or min.

$f' = (x, y) = (0, 0)$ are the candidates for maximum or minimum.

But how do we know maximum or minimum?

Answer is the second derivative tells us.

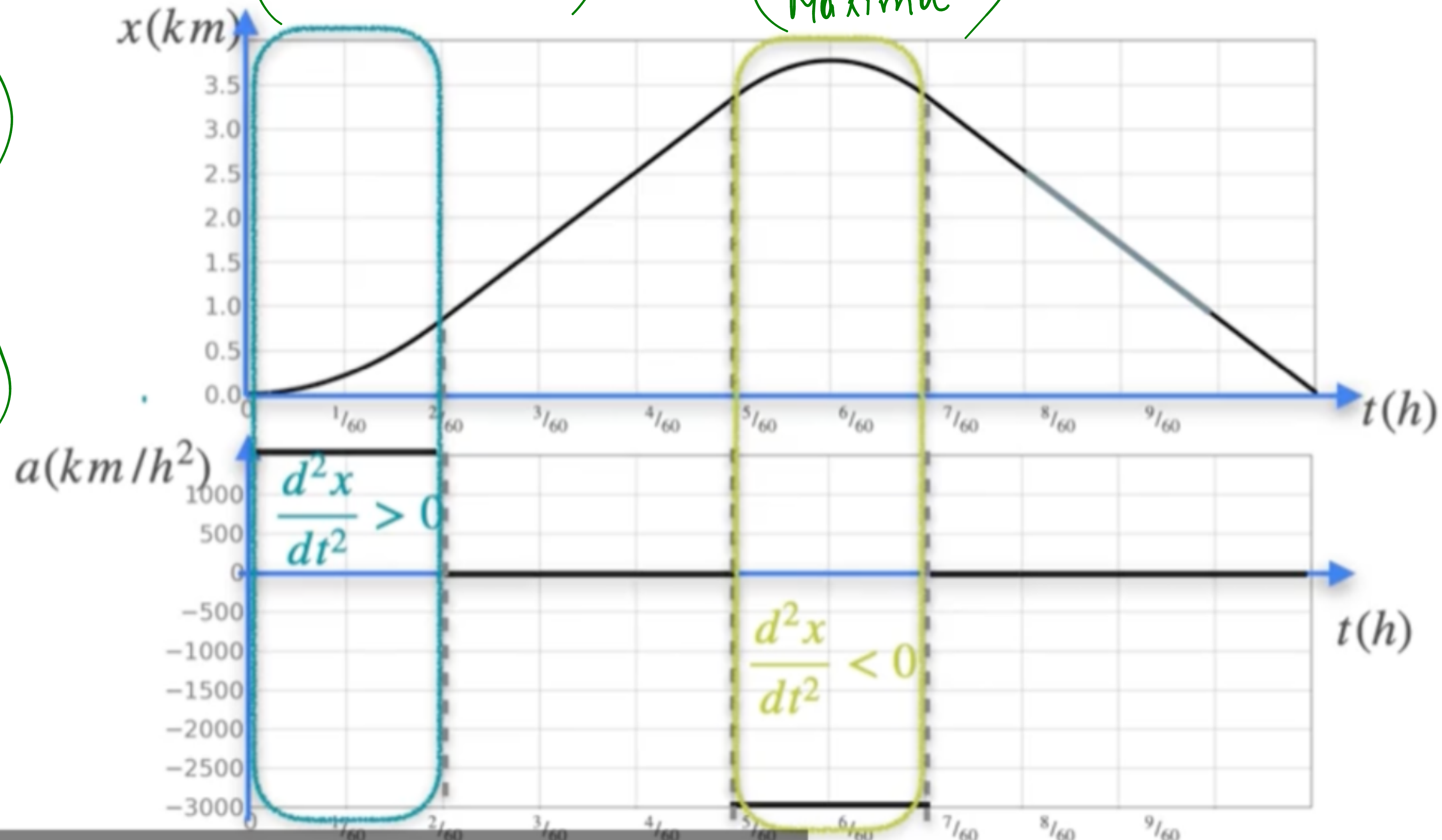
Second Derivative and Optimization

$$\frac{d^2x}{dt^2} > 0$$ (local minima)

$$\frac{d^2x}{dt^2} < 0$$ (local maxima)

$$\frac{d^2x}{dt^2} = 0$$ Inconclusive.

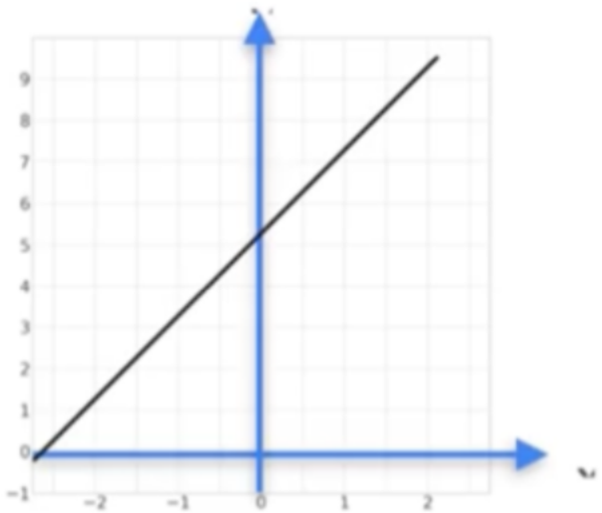
(Local Minima) (Local Maxima)



So the second derivative tells us, take a look at the first one.

Curvature

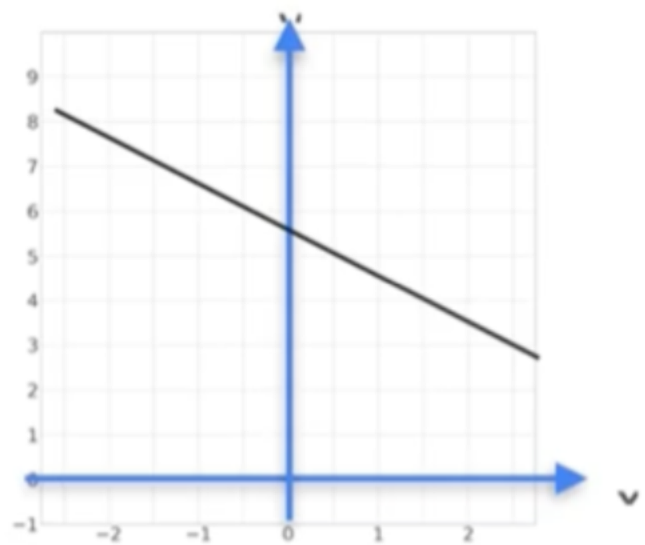
First derivative



Increasing

$f'(0) > 0$

↳ Maxima

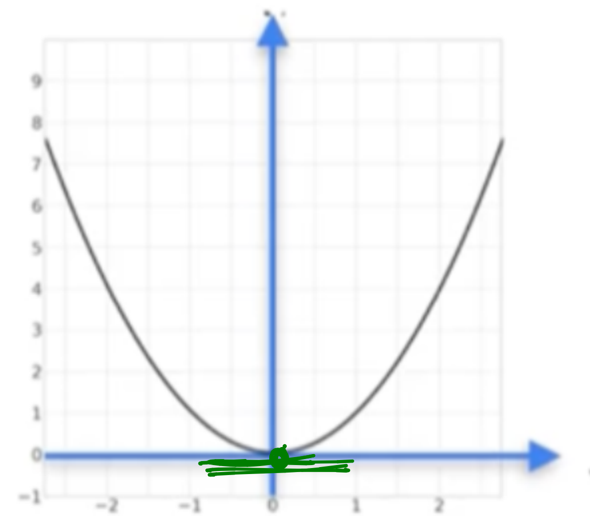


Decreasing

$f'(0) < 0$

↳ Minima

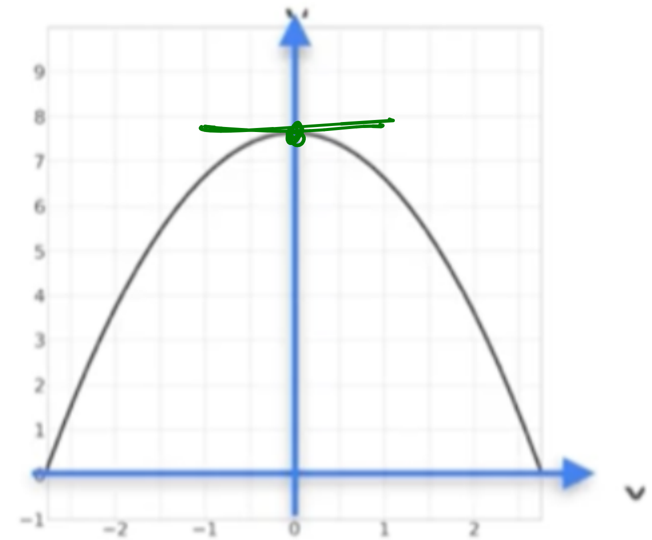
Second derivative



Concave up

$f''(0) > 0$

↳ Minima



Concave down

$f''(0) < 0$

↳ Maxima

DeepLearning.AI

Whereas when you have a concave down,

Hessian.

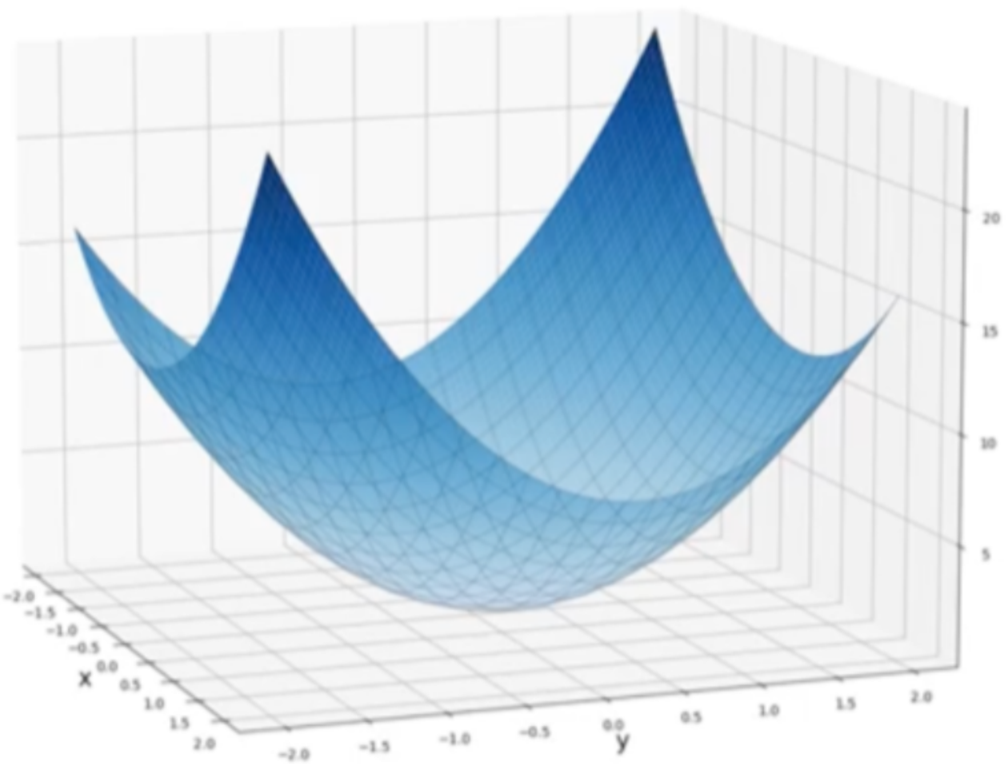
For multiple variables, the second derivative is actually a matrix full of second derivative called the Hessian.

Second Derivative

	1 variable	2 variables
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 of $f(x)$	???

That's the rate of change of the rate

Second Derivative



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

$$f(x,y) \xrightarrow{x} 4x - y \xrightarrow{x} 4 \quad f'_x(x,y) \text{ w.r.t } x$$

$$4x - y \xrightarrow{y} -1 \quad f'_x(x,y) \text{ w.r.t } y$$

$$f(x,y) \xrightarrow{y} 6y - x \xrightarrow{x} -1 \quad f'_y(x,y) \text{ w.r.t. } x$$

$$6y - x \xrightarrow{y} 6 \quad f'_y(x,y) \text{ w.r.t. } y$$

y) equals 2x squared plus
3y squared minus xy.

DeepLearning.AI

What Do These Mean?

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

Change in the change in the function w.r.t tiny changes in x and y

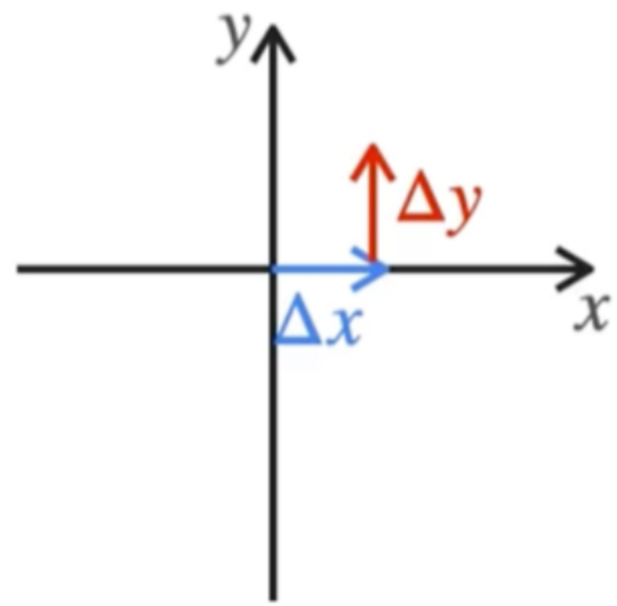
Same idea as with one variable!

Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis
2. They are the same!

(In most cases)



x of the derivative respect to y

DeepLearning.AI

Notation

	Leibniz's notation	Lagrange's Notation
Rate of change of $f'_x(x, y)$ w.r.t x	$\frac{\partial^2 f}{\partial x^2}$	$f_{xx}(x, y)$
Rate of change of $f'_y(x, y)$ w.r.t y	$\frac{\partial^2 f}{\partial y^2}$	$f_{yy}(x, y)$
Rate of change of $f'_x(x, y)$ w.r.t y	$\frac{\partial^2 f}{\partial x \partial y}$	$f_{xy}(x, y)$
Rate of change of $f'_y(x, y)$ w.r.t x	$\frac{\partial^2 f}{\partial y \partial x}$	$f_{yx}(x, y)$

as d squared f over dx squared,

DeepLearning.AI

Hessian Matrix

$$
\begin{tikzcd}
 & & & 4 \\
 & & 4x - y \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & & & -1 \\
f(x,y) = 2x^2 + 3y^2 - xy \arrow[uurr, "x"] \arrow[ddrr, "y"'] & & & \\
 & & & -1 \\
 & & 6y - x \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & & & 6
\end{tikzcd}
$$

$$\begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

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

$f'(x, y) = 2x + 2y$

$$
\begin{tikzcd}
 & 2x \arrow[r, "x"] \arrow[dr, "y"'] & 2 \\
 & & 0 \\
x^2 + y^2 \arrow[uur, "x"] \arrow[dr, "y"'] & & \\
 & 2y \arrow[r, "y"] \arrow[dr, "x"'] & 2 \\
 & & 0
\end{tikzcd}
$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

When we put these four together in a matrix, Hession Matrix keeps the f'' information whereas Jacobian Matrix f''.

Second Derivative

	1 variable		2 variables	
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 of $f(x)$	$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$	

Hession Matrix.

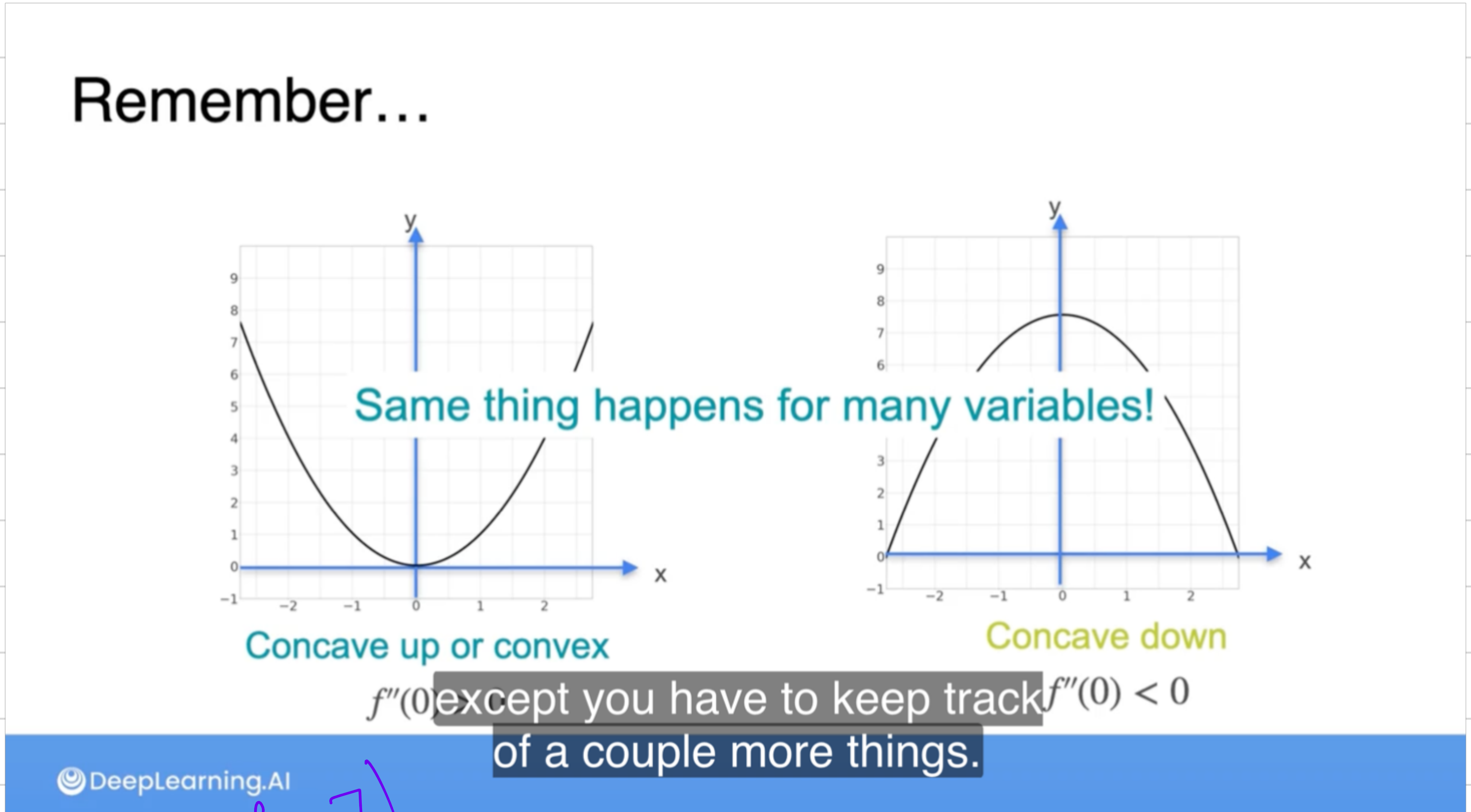
of all the second partial derivatives.

DeepLearning.AI

Concave up:

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

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$



Check if matrix +ve or -ve?

use → Eigen value.

$$\det(H(0,0) - \lambda I) = \det\left(\begin{bmatrix} 4-\lambda & -1 \\ -1 & 6-\lambda \end{bmatrix}\right)$$

$$\Rightarrow (4-\lambda)(6-\lambda) - (-1)(-1)$$

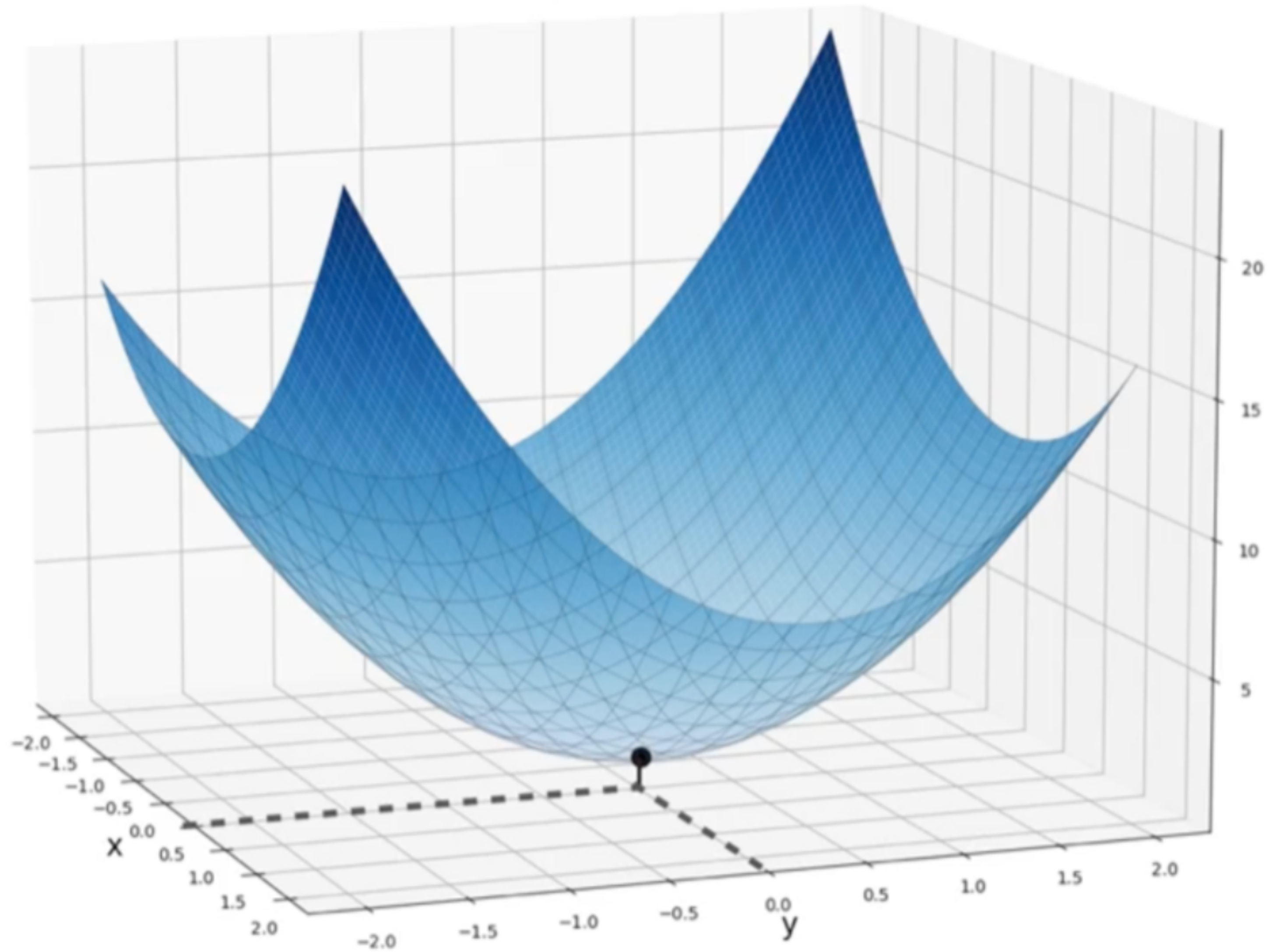
$$\Rightarrow \lambda^2 - 10\lambda + 23$$

$\rightarrow \lambda_1 = 6.41$ (+ve)

$\rightarrow \lambda_2 = 3.59$ (+ve)

Eigen values +ve so it is concave up → (0,0) is minimum.

Concave Up



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

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det\left(\begin{bmatrix} 4-\lambda & -1 \\ -1 & 6-\lambda \end{bmatrix}\right)$$

$$= (4-\lambda)(6-\lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23 \quad \boxed{\begin{matrix} \lambda_1 = 6.41 \\ \lambda_2 = 3.59 \end{matrix}}$$

> 0

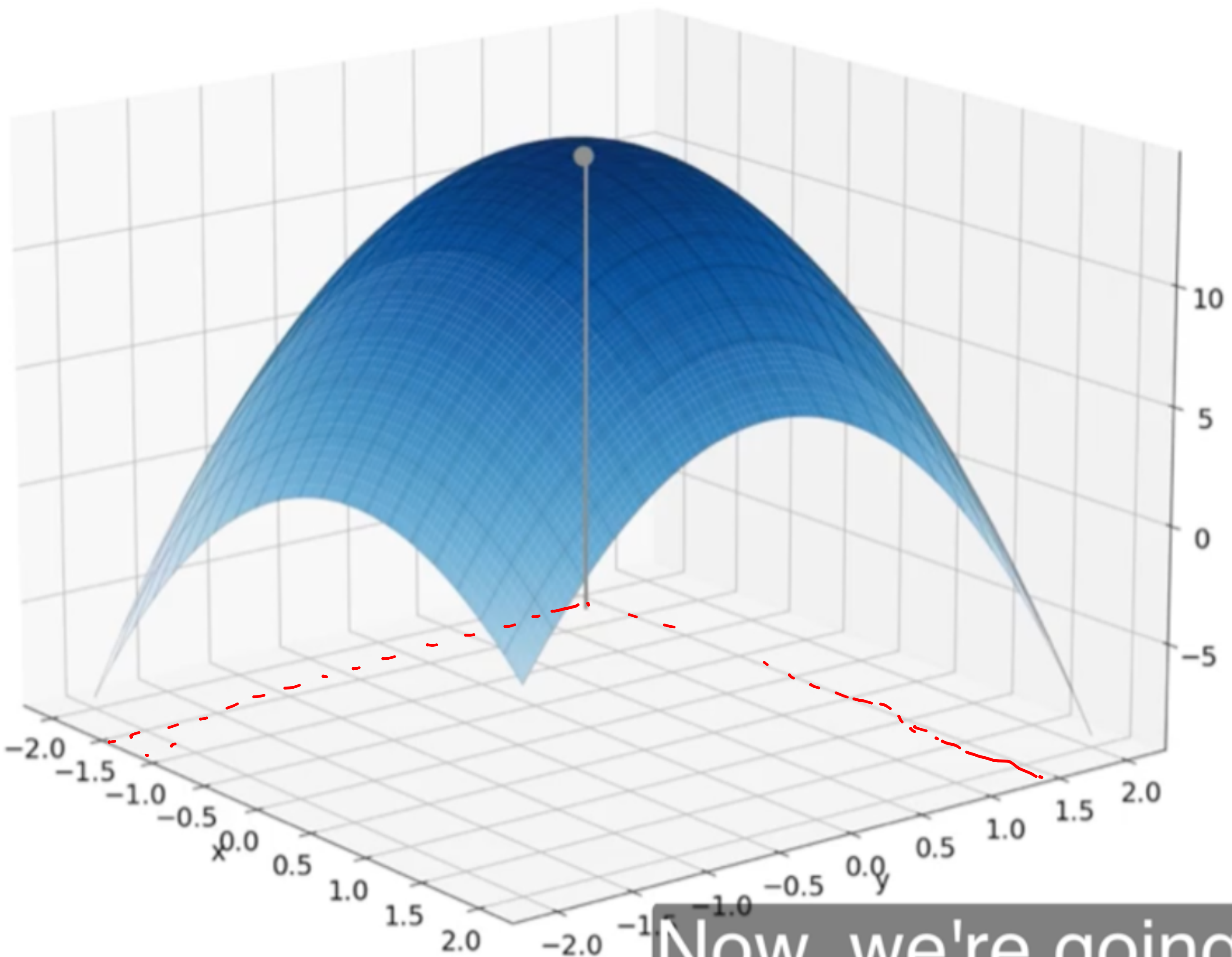
(0,0) is a minimum!

up and the 0.00 is a minimum.

DeepLearning.AI

Maximum point at $(0,0) = ?$

Concave Down



$$f(x, y) = -2x^2 + -3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = (-4-\lambda)(-6-\lambda) - (-1)(-1)$$

$$\Rightarrow \lambda^2 + 10\lambda + 23$$

$\lambda_1 \Rightarrow -3.59$

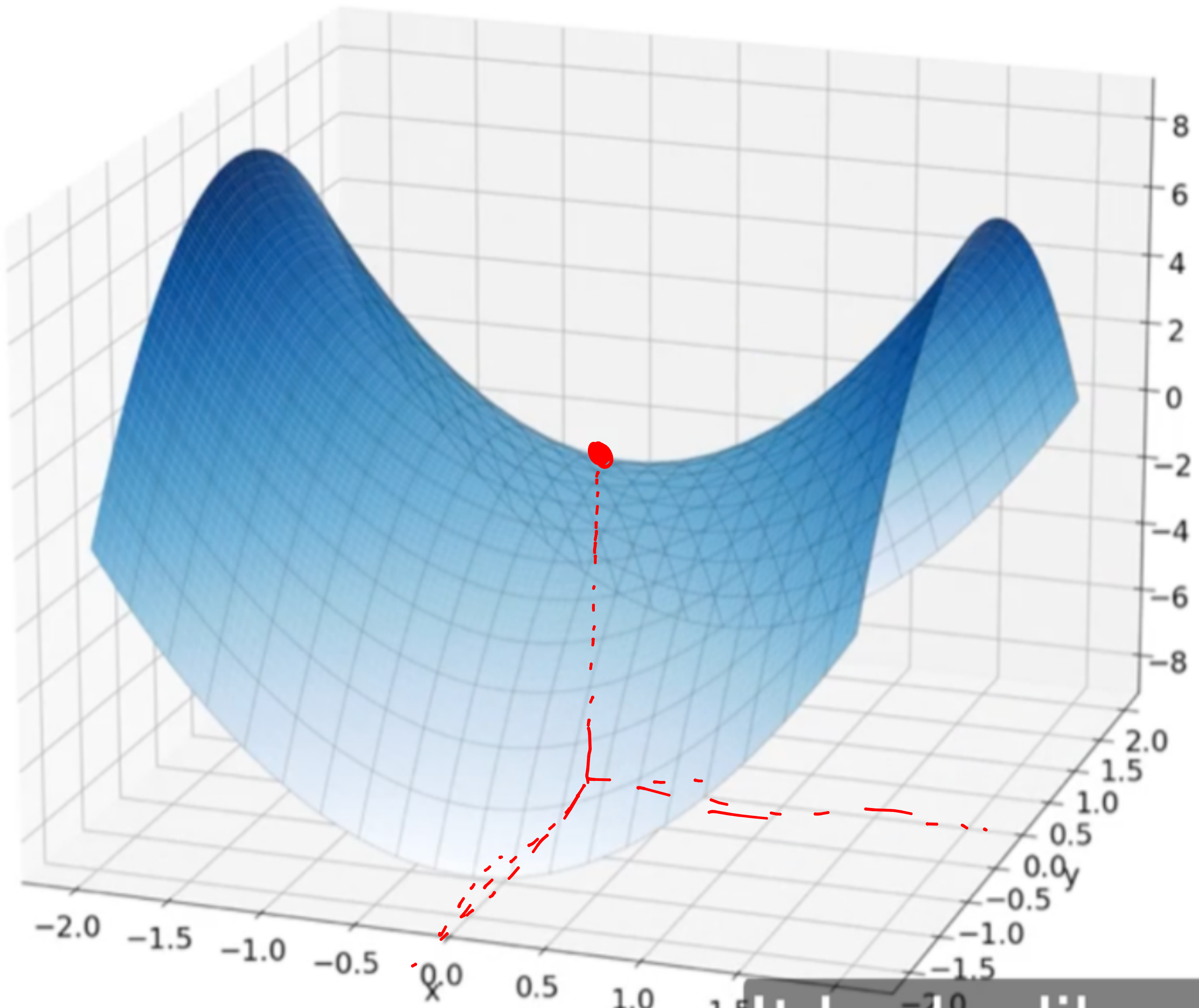
$\lambda_2 \Rightarrow -6.41$

Now, we're going to look at the Hessian matrix.

$(0,0)$ is a maximum! < 0

DeepLearning.AI

Saddle Point



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

$$\nabla f(x,y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = (4-\lambda)(-4-\lambda) - 0$$

$$\boxed{\begin{matrix} \lambda_1 = -4 \\ \lambda_2 = 4 \end{matrix}}$$

(0,0) is the saddle point

It looks like a saddle or a potato chip.

DeepLearning.AI

neither Max Nor Minimum

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \ldots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \,\&\, \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0 \,\&\, \lambda_2 < 0$	All $\lambda_i < 0$
Need more information	$f''(x) = 0$	Saddle point $\lambda_1 > 0 \,\&\, \lambda_2 < 0$ $\lambda_1 < 0 \,\&\, \lambda_2 > 0$ Or some $\lambda_i = 0$	Some $\lambda_i > 0$ and some $\lambda_j < 0$ OR At least one $\lambda_i = 0$

you have a local minimum.

DeepLearning.AI

Newton's Method:

1 variable

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

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \underbrace{H^{-1}(x_k, y_k)}_{2\times 2} \, \underbrace{\nabla f(x_k, y_k)}_{2\times 1}$$

When working with 2 variables the order matters.

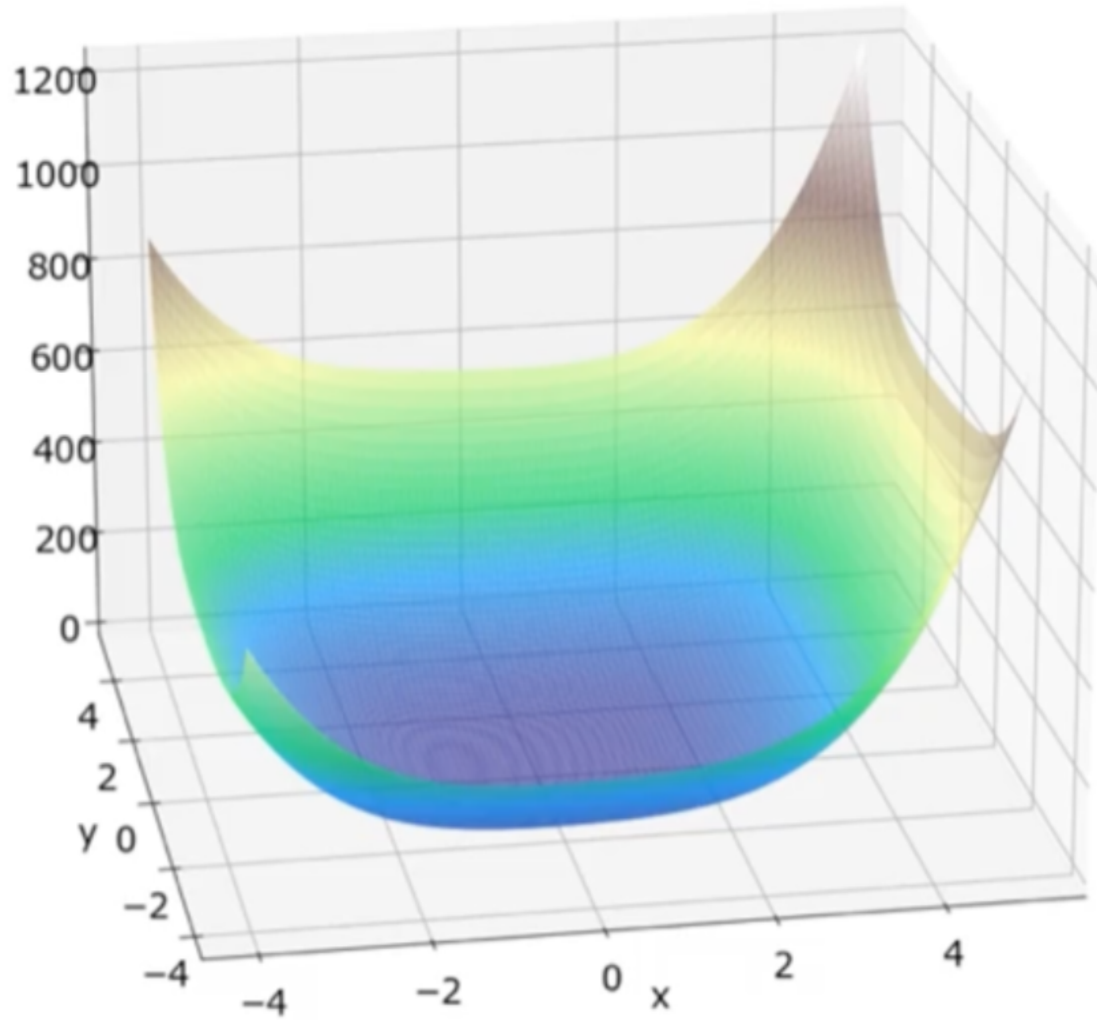
Gradient.

$$\nabla f(x,y) = \begin{bmatrix} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{bmatrix}$$

Hessian

$$H(x,y) = \begin{bmatrix} 12x^2 + 8 - 0.4 & -1-0.4x \\ -1-0.4x & 9.6y^2+4 \end{bmatrix}$$

An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$f(x,y) \xrightarrow{x} 4x^3 + 8x - y - 0.4xy$$

$$4x^3 + 8x - y - 0.4xy \xrightarrow{x} 12x^2 + 8 - 0.4y$$

$$4x^3 + 8x - y - 0.4xy \xrightarrow{y} -1 - 0.4x$$

$$f(x,y) \xrightarrow{y} 3.2y^3 + 4y - x - 0.2x^2$$

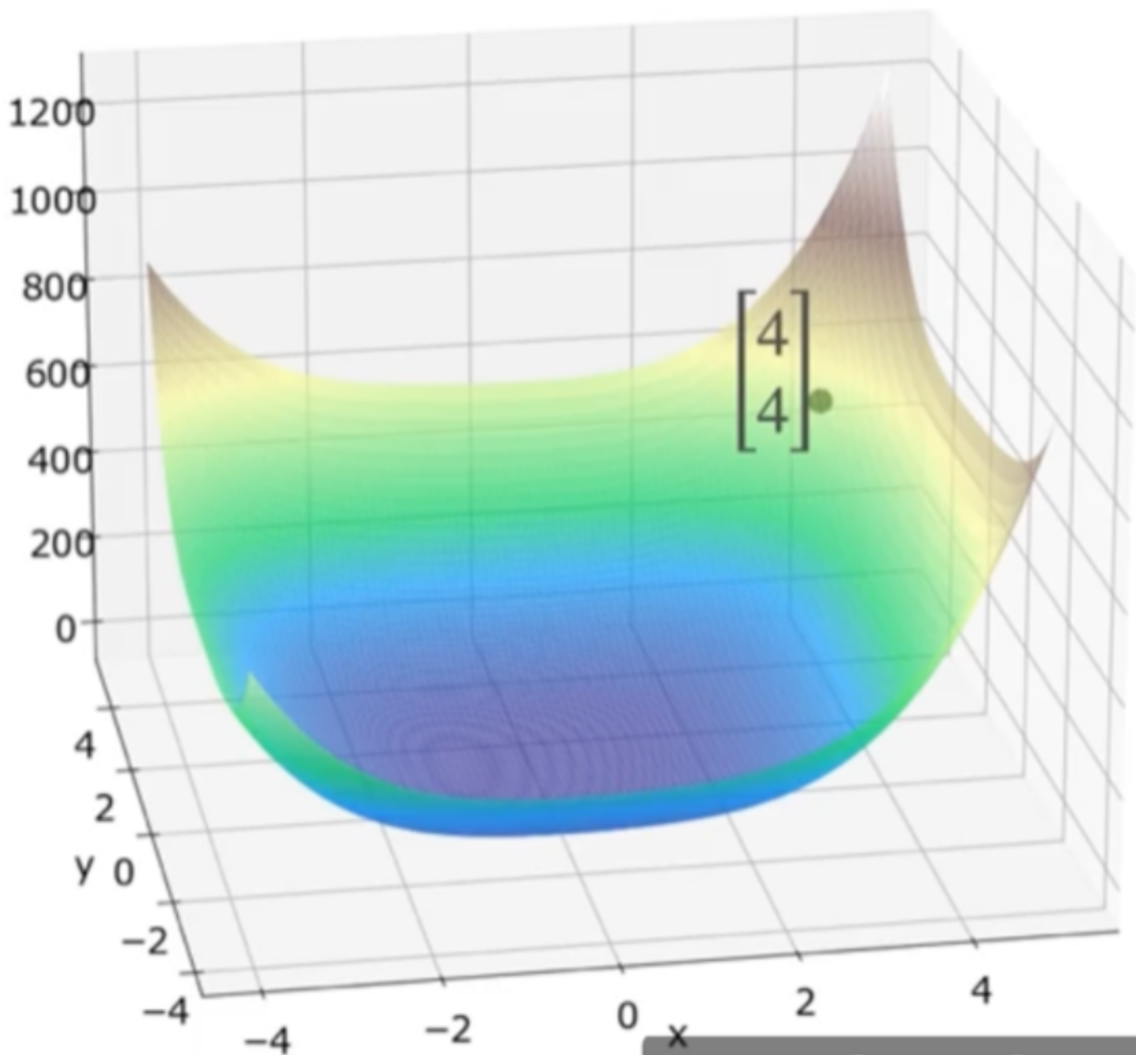
$$3.2y^3 + 4y - x - 0.2x^2 \xrightarrow{x} -1 - 0.4x$$

$$3.2y^3 + 4y - x - 0.2x^2 \xrightarrow{y} 9.6y^2 + 4$$

So first,
let's take a look at the two partial

DeepLearning.AI

An Example



Start at some point $\begin{bmatrix}x_0\\y_0\end{bmatrix} = \begin{bmatrix}4\\4\end{bmatrix}$

$$\nabla f(4,4) = \begin{bmatrix}277.6\\213.6\end{bmatrix}$$

$$H(4,4) = \begin{bmatrix}198.4 & -2.6\\-2.6 & 157.6\end{bmatrix}$$

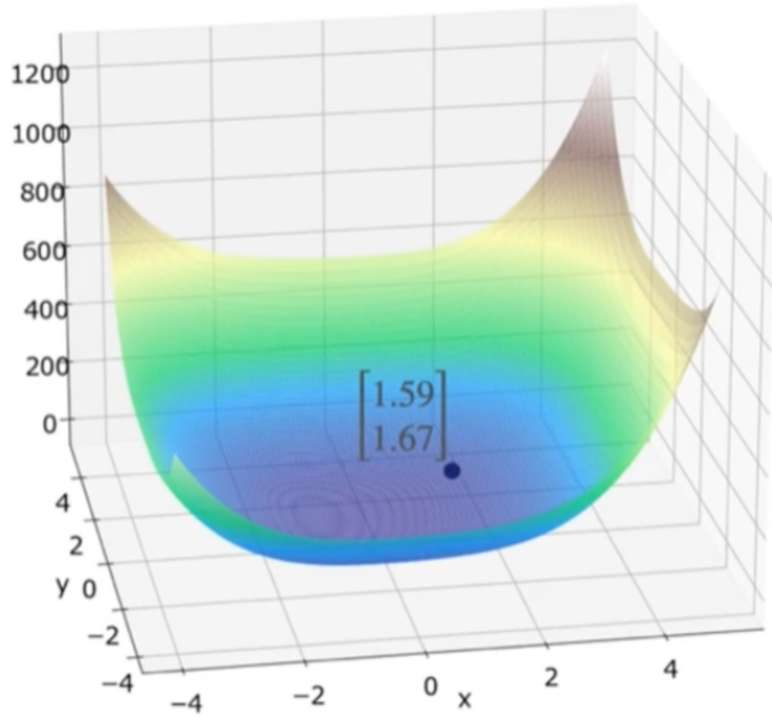
$$\begin{bmatrix}x_1\\y_1\end{bmatrix} = \begin{bmatrix}4\\4\end{bmatrix} - \begin{bmatrix}198.4 & -2.6\\-2.6 & 157.6\end{bmatrix}^{-1} \begin{bmatrix}277.6\\213.6\end{bmatrix}$$

$$\begin{bmatrix}x_1\\y_1\end{bmatrix} \Rightarrow \begin{bmatrix}2.58\\2.62\end{bmatrix}$$

we simply put the numbers 4 and 4,
and now let's find the Hessian.

DeepLearning.AI

An Example



$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} \quad H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} - \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}^{-1} \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

$$= \begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$

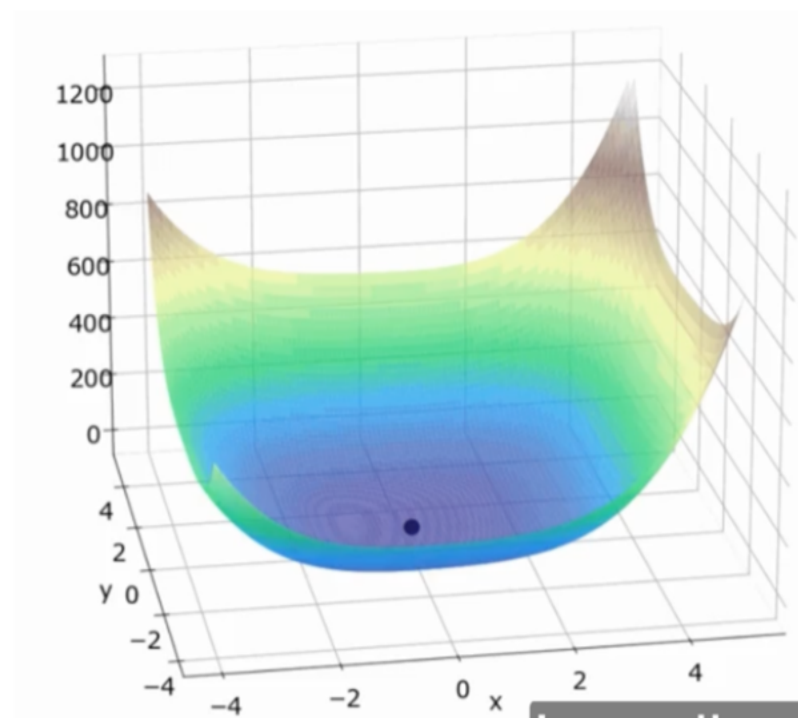
DeepLearning.AI

So we get the 1.59, 1.6

Newton's method to optimize functions of many variables.

Newton's method for two variables is also faster - able to approximate closely the 0 of f.

An Example



Repeat until you are close enough to the actual zero!

Needed $k = 8$ steps

$$\begin{bmatrix} x_8 \\ y_8 \end{bmatrix} = \begin{bmatrix} 4.15\ 10^{-17} \\ -2.05\ 10^{-17} \end{bmatrix}$$

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Is really, really, really close to 0,
0, and 0, 0 is actually optimal

DeepLearning.AI