MATHEMATICS OF MACHINE LEARNING

AGENDA

- Vectors
- Matrices
- Derivatives

Why vectors and matrices?

- Most common form of data organization for machine learning is a 2D array, where
  - *rows* represent samples (records, items, datapoints)
  - *columns* represent attributes (features, variables)
- Natural to think of each sample as a *vector* of attributes, and whole array as a *matrix*

vector

Refund	Marital Status	Taxable Income	Cheat
Yes	Single	125K	No
No	Married	100K	No
No	Single	70K	No
Yes	Married	120K	No
No	Divorced	95K	Yes
No	Married	60K	No
Yes	Divorced	220K	No
No	Single	85K	Yes
No	Married	75K	No
No	Single	90K	Yes

matrix

Vectors

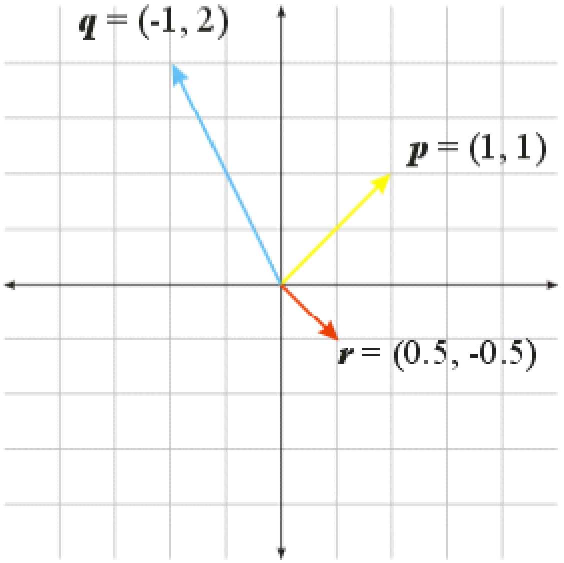
- Definition: an *n*-tuple of values (usually real numbers).
  - *n* referred to as the *dimension* of the vector
  - *n* can be any positive integer, from 1 to infinity
- Can be written in column form or row form
  - Column form is conventional
  - Vector elements referenced by subscript

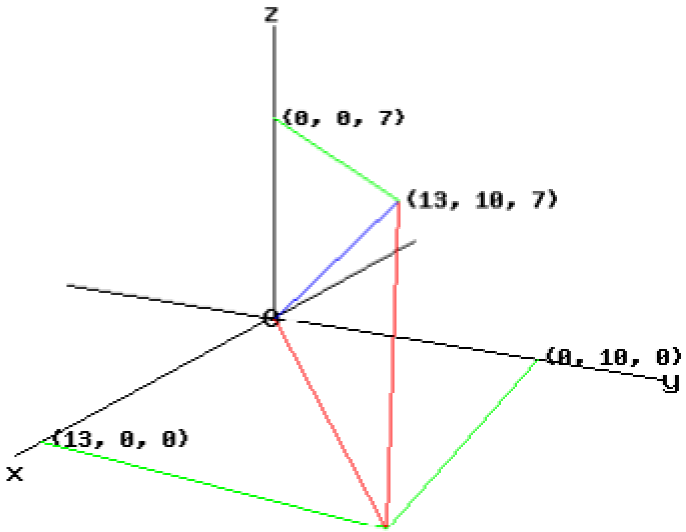
$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbf{x}^{\mathrm{T}} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$$

$^{\mathrm{T}}$ means "transpose"

Vectors

- Can think of a vector as:
  - a point in space *or*
  - a directed line segment with a magnitude and direction

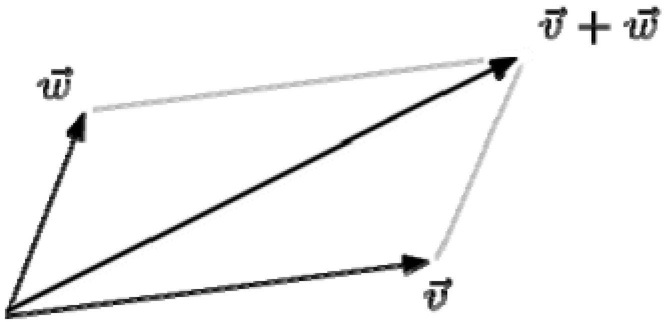




Vector arithmetic

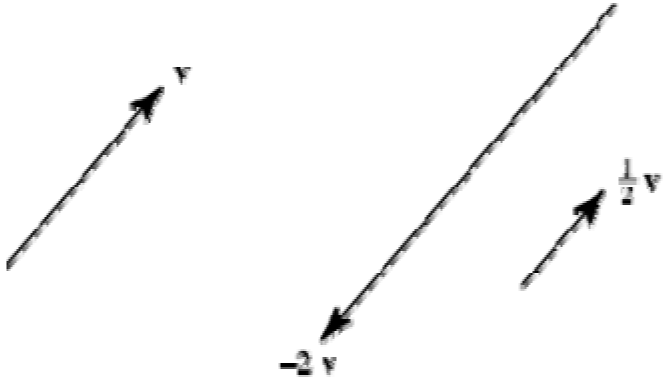
- Addition of two vectors
  - add corresponding elements

    $$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 & \cdots & x_n + y_n \end{pmatrix}^{\mathrm{T}}$$
  - result is a vector



- Scalar multiplication of a vector
  - multiply each element by scalar

    $$\mathbf{y} = a\mathbf{x} = \begin{pmatrix} a x_1 & \cdots & a x_n \end{pmatrix}^{\mathrm{T}}$$
  - result is a vector



Vector arithmetic

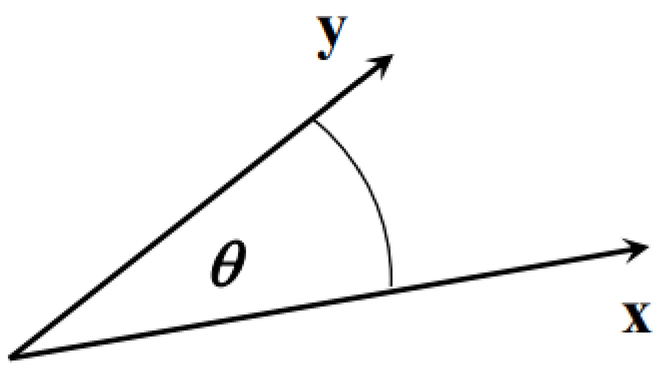
- Dot product of two vectors
  - multiply corresponding elements, then add products

$$a = \mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

  - result is a <u>scalar</u>

- Dot product alternative form

$$a = \mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta)$$



Matrices

- Definition: an *m* x *n* two-dimensional array of values (usually real numbers).
  - *m* rows
  - *n* columns
- Matrix referenced by two-element subscript
  - first element in subscript is row
  - second element in subscript is column
  - example: $\mathbf{A}_{24}$ or a_{24} is element in second row, fourth column of **A**

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

Matrices

- A vector can be regarded as special case of a matrix, where one of matrix dimensions = 1.
- Matrix *transpose* (denoted ^{T})
  - swap columns and rows
    - row 1 becomes column 1, etc.
  - *m* x *n* matrix becomes *n* x *m* matrix
  - example:

$$\mathbf{A} = \begin{pmatrix} 2 & 7 & -1 & 0 & 3 \\ 4 & 6 & -3 & 1 & 8 \end{pmatrix} \qquad \mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 2 & 4 \\ 7 & 6 \\ -1 & -3 \\ 0 & 1 \\ 3 & 8 \end{pmatrix}$$

Matrix arithmetic

- Addition of two matrices
  - matrices must be same size
  - add corresponding elements:

$$c_{ij} = a_{ij} + b_{ij}$$

  - result is a matrix of same size

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

- Scalar multiplication of a matrix
  - multiply each element by scalar:

$$b_{ij} = d \cdot a_{ij}$$

  - result is a matrix of same size

$$\mathbf{B} = d \cdot \mathbf{A} = \begin{pmatrix} d \cdot a_{11} & \cdots & d \cdot a_{1n} \\ \vdots & \ddots & \vdots \\ d \cdot a_{m1} & \cdots & d \cdot a_{mn} \end{pmatrix}$$

Matrix arithmetic

- Matrix-matrix multiplication
  - vector-matrix multiplication just a special case

TO THE BOARD!!

- Multiplication is associative

$$\mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C}) = (\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{C}$$

- Multiplication is *not* commutative

$$\mathbf{A} \cdot \mathbf{B} \neq \mathbf{B} \cdot \mathbf{A} \quad \text{(generally)}$$

- Transposition rule:

$$(\mathbf{A} \cdot \mathbf{B})^T = \mathbf{B}^T \cdot \mathbf{A}^T$$

Matrix arithmetic

- *RULE*: In any chain of matrix multiplications, the *column* dimension of one matrix in the chain must match the *row* dimension of the *following* matrix in the chain.
- Examples

A 3 x 5 **B** 5 x 5 **C** 3 x 1

Right:

$\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{A}^T$ $\mathbf{C}^T \cdot \mathbf{A} \cdot \mathbf{B}$ $\mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{B}$ $\mathbf{C} \cdot \mathbf{C}^T \cdot \mathbf{A}$

Wrong:

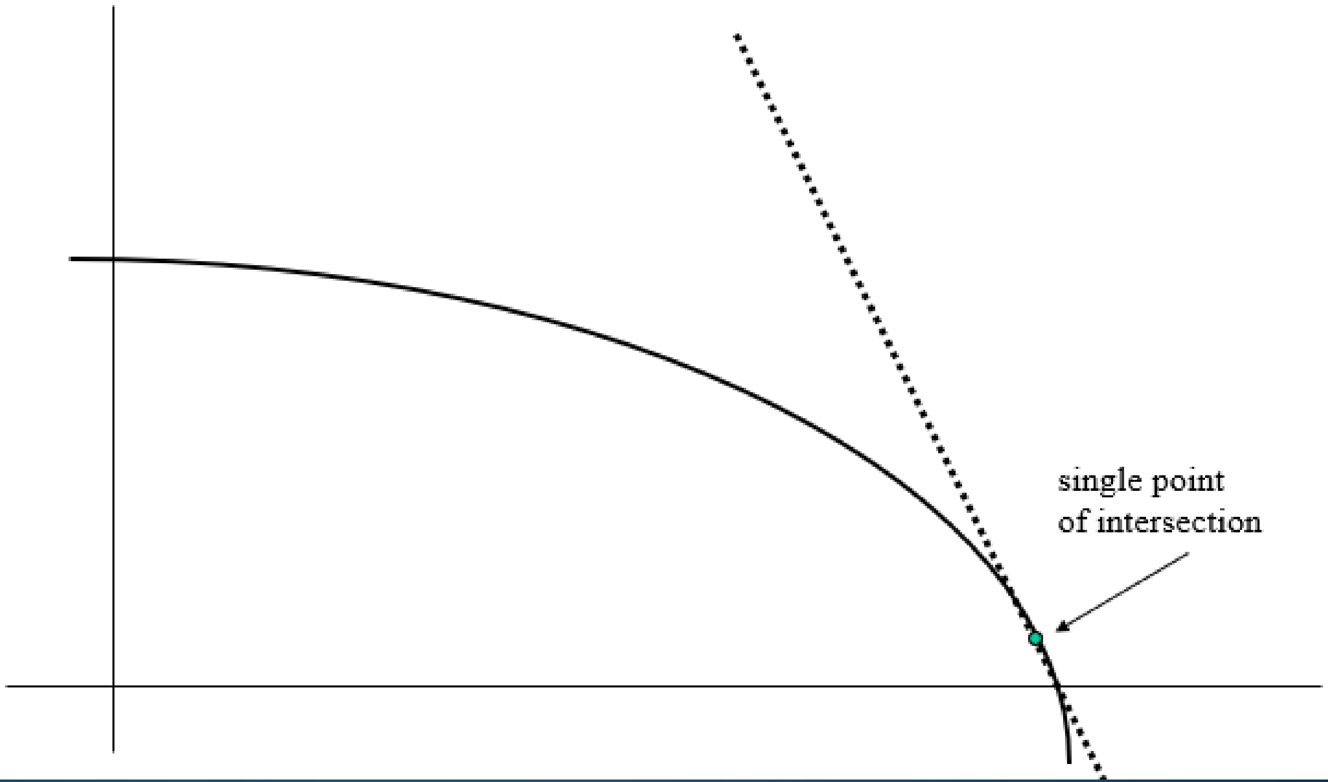
$\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{A}$ $\mathbf{C} \cdot \mathbf{A} \cdot \mathbf{B}$ $\mathbf{A} \cdot \mathbf{A}^T \cdot \mathbf{B}$ $\mathbf{C}^T \cdot \mathbf{C} \cdot \mathbf{A}$

DERIVATIVES

What is a *derivative*?

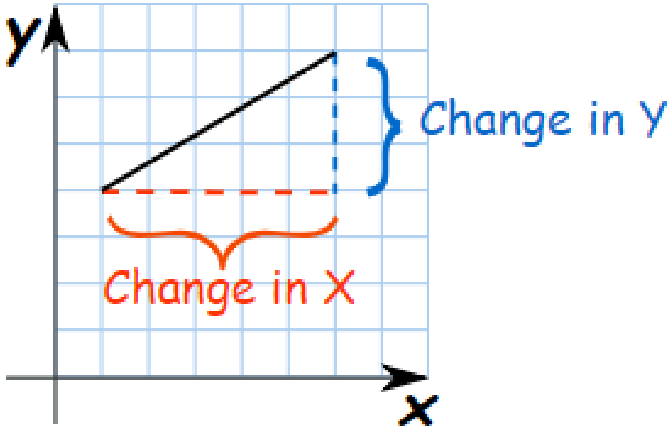
- A function
- the rate of change of a function
- the slope of the line tangent to the curve

The tangent line

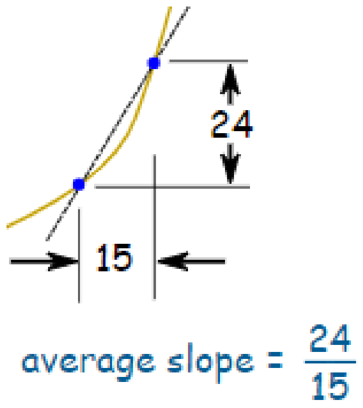


It is all about slope!

$$\text{Slope} = \frac{\text{Change in Y}}{\text{Change in X}}$$

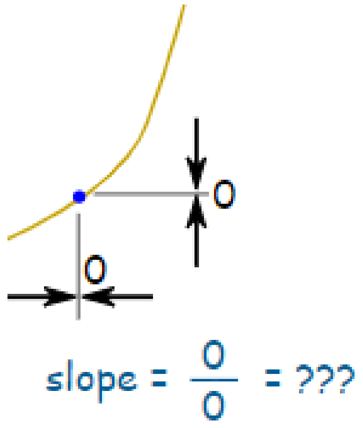


We can find an **average** slope between two points.



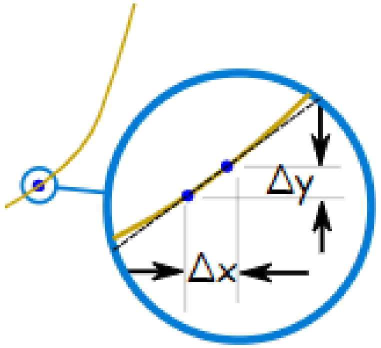
But how do we find the slope **at a point**?

There is nothing to measure!



But with derivatives we use a small difference ...

... then have it **shrink towards zero**.

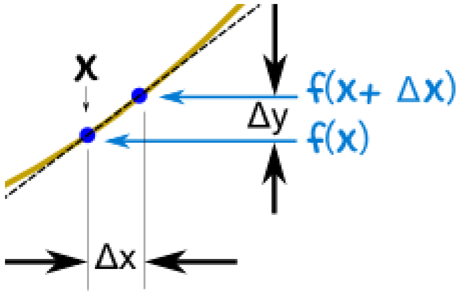


Let us Find a Derivative!

To find the derivative of a function $y = f(x)$ we use the slope formula:

$$\text{Slope} = \frac{\text{Change in Y}}{\text{Change in X}} = \frac{\Delta y}{\Delta x}$$

And (from the diagram) we see that:



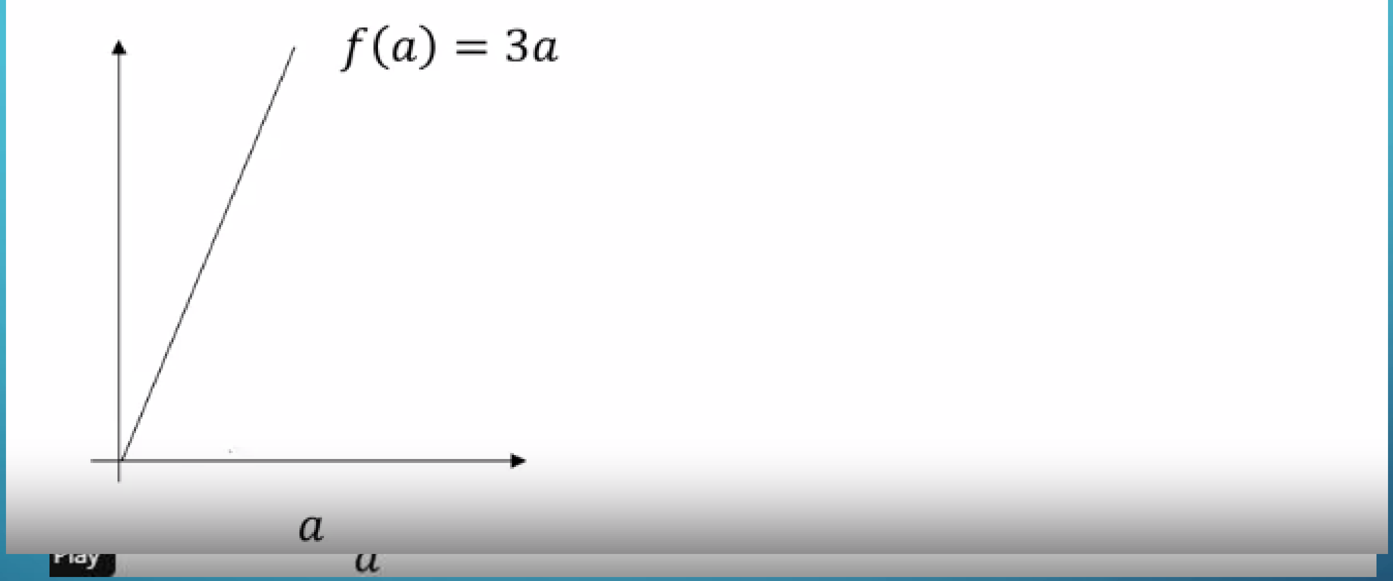
x changes from x to $x+\Delta x$

y changes from $f(x)$ to $f(x+\Delta x)$

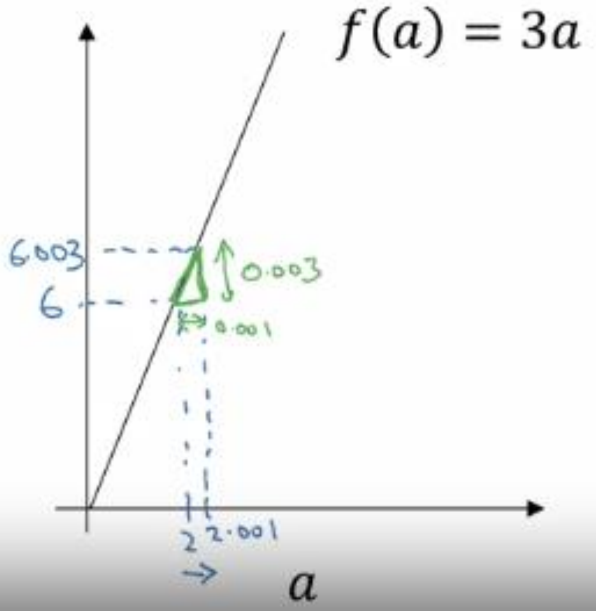
Now follow these steps:

- Fill in this slope formula: $\frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}$
- Simplify it as best we can
- Then make **Δx** shrink towards zero.

Intuition about derivatives



Intuition about derivatives

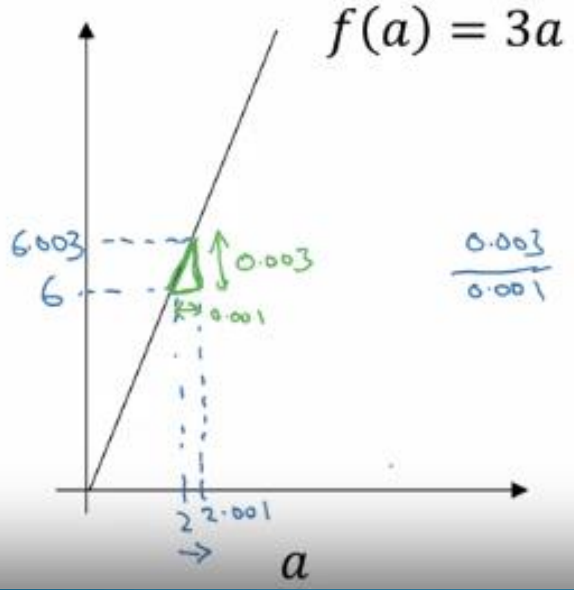


$a = 2 \qquad f(a) = 6$

$a = 2.001 \qquad f(a) = 6.003$

slope (derivative) of $f(a)$
at $a=2$ is 3

Intuition about derivatives

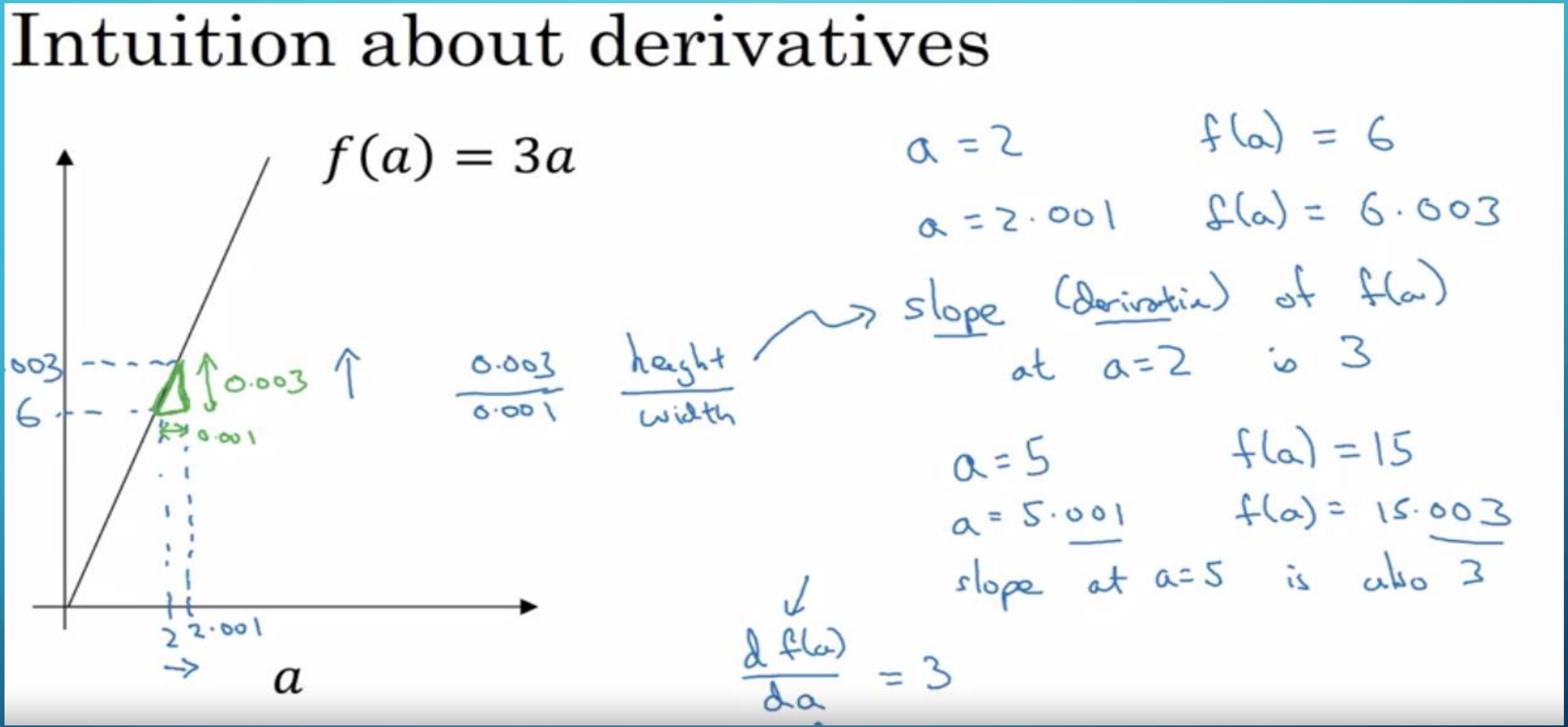


$\frac{0.003}{0.001}$ $\frac{\text{height}}{\text{width}}$ → slope (derivative) of $f(a)$ at $a=2$ is 3

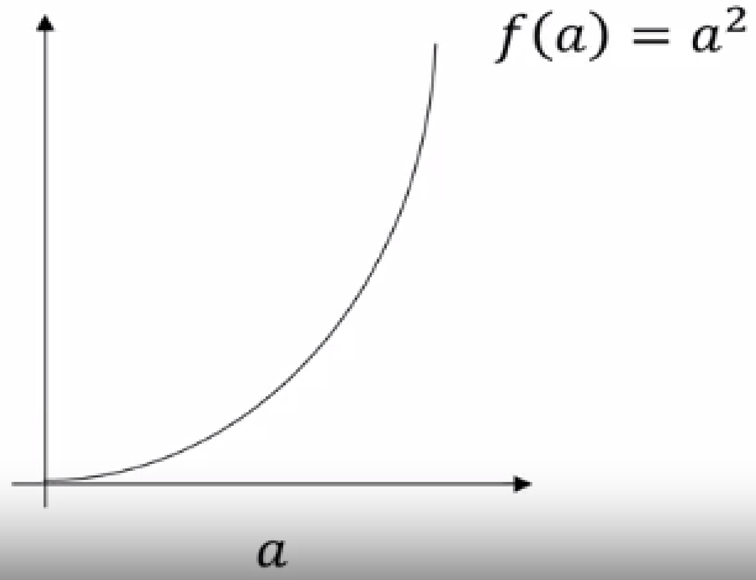
$a = 2$ $f(a) = 6$

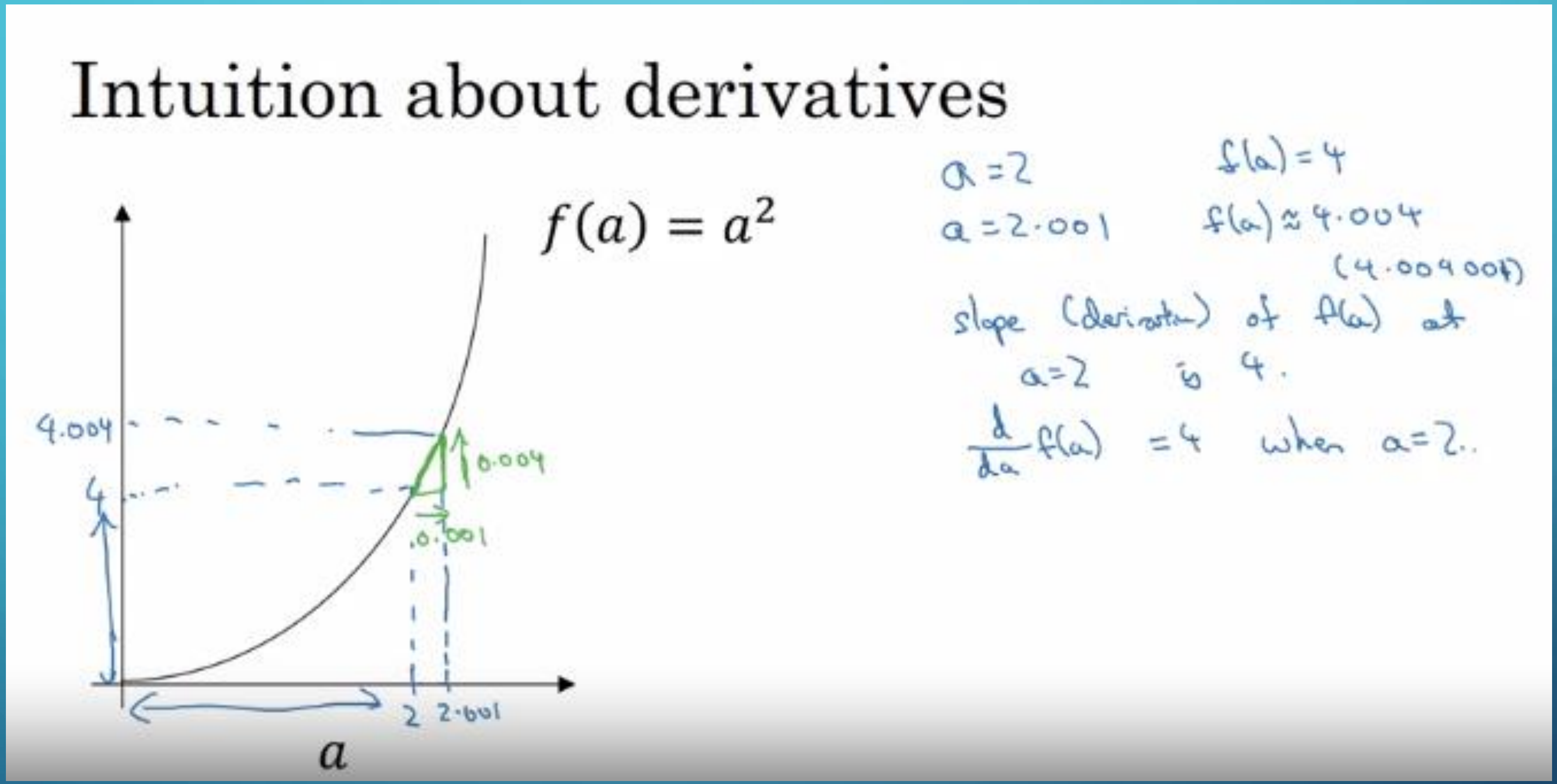
$a = 2.001$ $f(a) = 6.003$

- Similarly for a=5 …

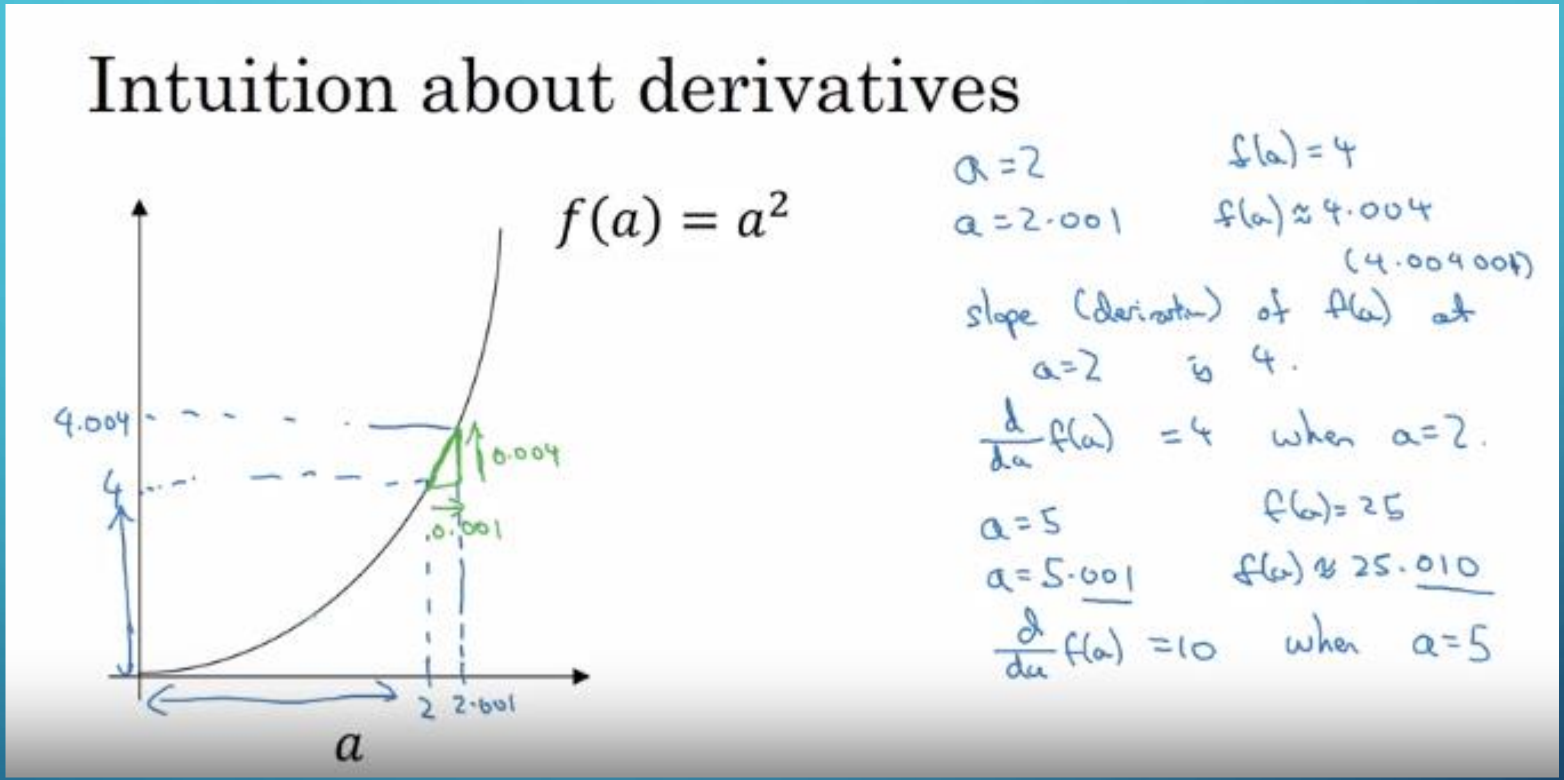


Intuition about derivatives



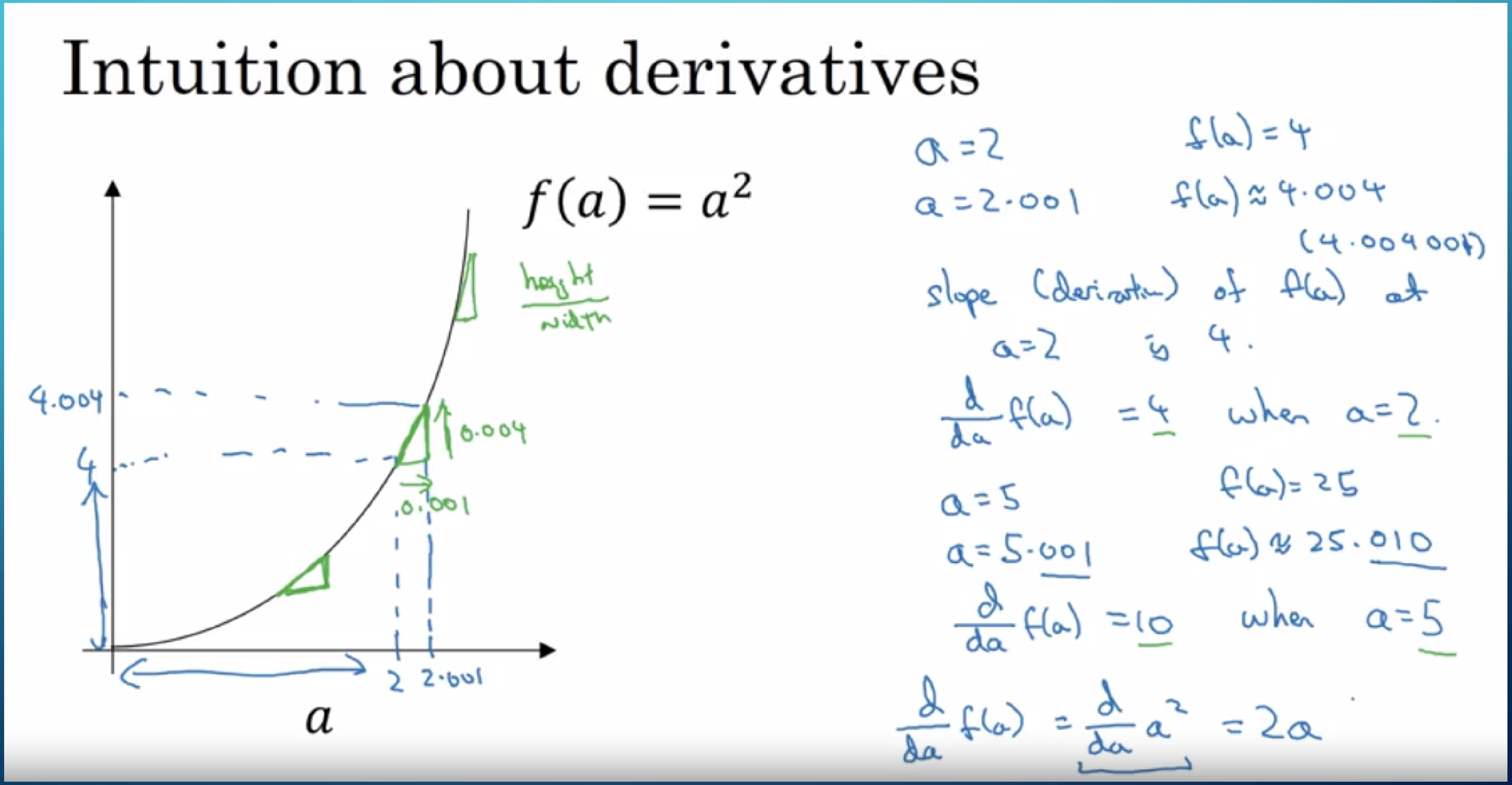

Intuition about derivatives
f(a) = a²
a = 2   f(a) = 4
a = 2.001   f(a) ≈ 4.004
(4.004001)
slope (derivative) of f(a) at
a=2 is 4.
d/da f(a) = 4 when a=2..
4.004
4
0.004
0.001
2
2.001
a

- Now for a=5:



Intuition about derivatives
$f(a) = a^2$
4.004
4
0.004
0.001
2
2.001
a
a=2 f(a)=4
a=2.001 f(a) ≈ 4.004
(4.004001)
slope (derivative) of f(a) at a=2 is 4.
$\frac{d}{da} f(a) = 4$ when a=2.
a=5 f(a)=25
a=5.001 f(a) ≈ 25.010
$\frac{d}{da} f(a) = 10$ when a=5

- Trend of derivatives show that slope depends on the value of a.
- Slope is always twice of a.



Example: the function $\mathbf{f(x) = x^2}$

We know $\mathbf{f(x)} = \mathbf{x^2}$, and we can calculate $\mathbf{f(x+\Delta x)}$:

Start with: $\mathbf{f(x+\Delta x) = (x+\Delta x)^2}$

Expand $(x + \Delta x)^2$: $\mathbf{f(x+\Delta x) = x^2 + 2x\,\Delta x + (\Delta x)^2}$

The slope formula is:	$\dfrac{f(x+\Delta x) - f(x)}{\Delta x}$
Put in $\mathbf{f(x+\Delta x)}$ and $\mathbf{f(x)}$:	$\dfrac{x^2 + 2x\,\Delta x + (\Delta x)^2 - x^2}{\Delta x}$
Simplify (x^2 and $-x^2$ cancel):	$\dfrac{2x\,\Delta x + (\Delta x)^2}{\Delta x}$
Simplify more (divide through by Δx):	$= 2x + \Delta x$
Then **as Δx heads towards 0** we get:	$= 2x$

Result: the derivative of $\mathbf{x^2}$ is $\mathbf{2x}$

THANK YOU