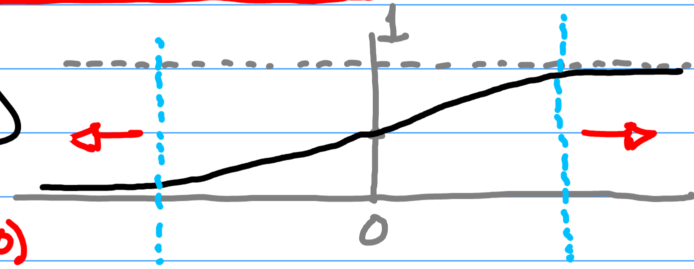
SOME HANDCRAFTED NN EXAMPLES

$$\boxed{h(a) = \frac{1}{1+e^{-a}}} \quad \text{(sigmoid)}$$



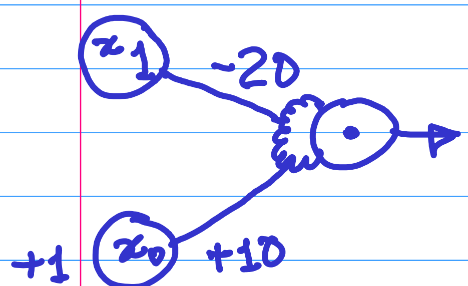
(0) region -4.6 $+4.6$ (1) region

$$e^{4.6} \approx 100$$

when $a = -4.6$, $h(a) = \frac{1}{1+e^{4.6}} \approx \frac{1}{1+100} \approx 0.01$

when $a = +4.6$, $h(a) = \frac{1}{1+e^{-4.6}} \approx \frac{1}{1+1/100} = \frac{100}{101} \approx 0.99$

Example 1 : NOT



$$a = \underline{w}^T\underline{x} + b$$
$$= -20x_1 + 10$$

x_1	$\overline{x_1}$
0	1
1	0

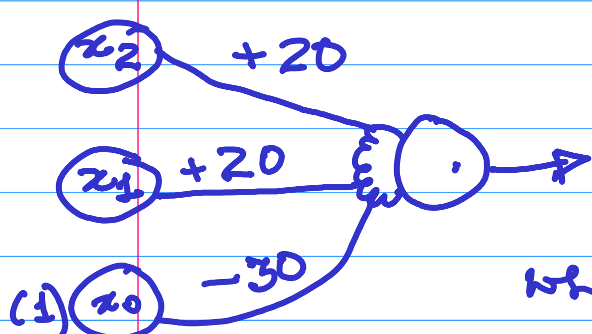
$x_1 = 0$, $a = 10$, $h(+10) \approx 1$

$x_1 = 1$, $a = -10$, $h(-10) \approx 0$

Example 2 : AND

$$a = \underline{w}^T\underline{x} + b$$

x_2	x_1	$x_2 \cdot x_1$
0	0	0
0	1	0
1	0	0
1	1	1



$$a = 20x_2 + 20x_1 - 30$$

when $x_2 = 0$, $x_1 = 0$: $a = -30$

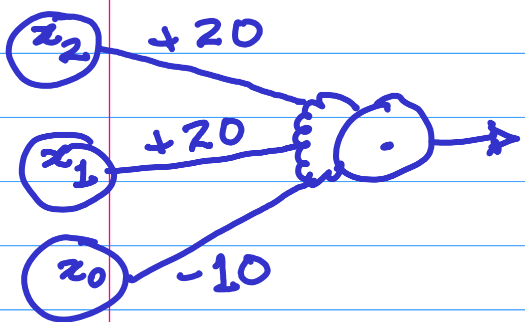
$h(-30) \approx 0$

when $x_2 = 0$, $x_1 = 1$, $a = 20(0) + 20(1) - 30 = -10$, $h(-10) \approx 0$

when $x_2 = 1$, $x_1 = 0$, $a = 20(1) + 20(0) - 30 = -10$, $h(-10) \approx 0$

when $x_2 = 1$, $x_1 = 1$, $a = 20(1) + 20(1) - 30 = +10$, $h(+10) \approx 1$

$\boxed{\text{Example 3 : OR}}$



x_2	x_1	$x_2 + x_1$
0	0	0
0	1	1
1	0	1
1	1	1

$$a = \underline{w}^T \underline{x} + b$$

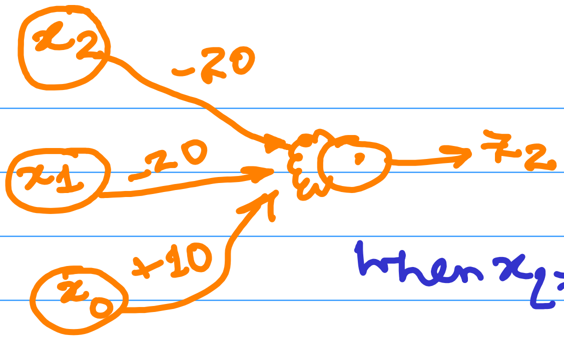
$$a = 20x_2 + 20x_1 - 10$$

when $x_2 = 0$, $x_1 = 0$: $a = 20(0) + 20(0) - 10 = -10$, $h(-10) \approx 0$

when $x_2 = 0$, $x_1 = 1$: $a = 20(0) + 20(1) - 10 = +10$, $h(+10) \approx 1$

when $x_2 = 1$, $x_1 = 0$: $a = 20(1) + 20(0) - 10 = +10$, $h(+10) \approx 1$

when $x_2 = 1$, $x_1 = 1$: $a = 20(1) + 20(1) - 10 = +30$, $h(+30) \approx 1$



$a = \underline{w}^T \underline{x} + b$

$= -20x_2 - 20x_1 + 10$

when $x_2 = 0, x_1 = 0$: $a = -20(0) - 20(0) + 10 = 10$

$h(10) \approx 1$

when $x_2 = 0, x_1 = 1$: $a = -20(0) - 20(1) + 10 = -10$

$h(-10) \approx 0$

when $x_2 = 1, x_1 = 0$: $a = -20(1) - 20(0) + 10 = -10$

$h(-10) \approx 0$

when $x_2 = 1, x_1 = 1$: $a = -20(1) - 20(1) + 10 = -30$

$h(-30) \approx 0$

$z_2 = \overline{x_2} \cdot \overline{x_1}$

$y = \overline{x_2} \cdot \overline{x_1} + x_2 . x_1 = x_2 \odot x_1$.

(*) This is an informal (non-mathematical, intuitive) manifestation of a fundamental notion →

"A feedforward NN with one hidden layer can represent any Boolean function"

(*) This also gives an intuitive (non-mathematical) perspective of a more general result: →

"Multi-layer feedforward NNs with non-linear activation functions are universal approximators — they can approximate any function arbitrarily well."

Building Block (Input: 2-D: Image)

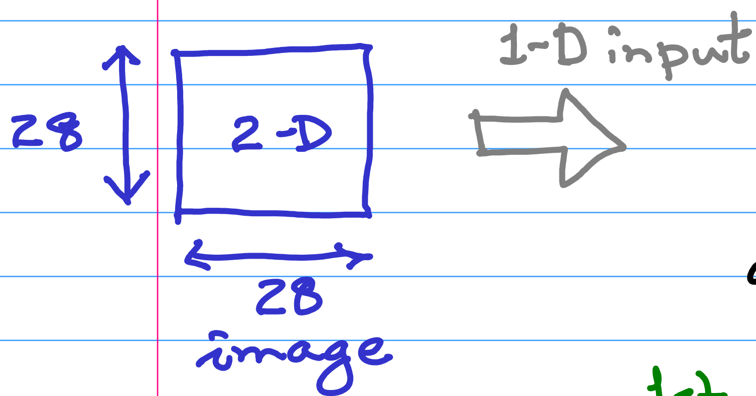
MNIST Numeral database: 28×28 images
(greyscale: not binary (not 0 and 255, or normalised 0 and 1))
~smooth

→ shades of grey as well, though most of the image is black (0) or white (255, or 1, normalised).

Images of the 10 numerals: 0 to 9

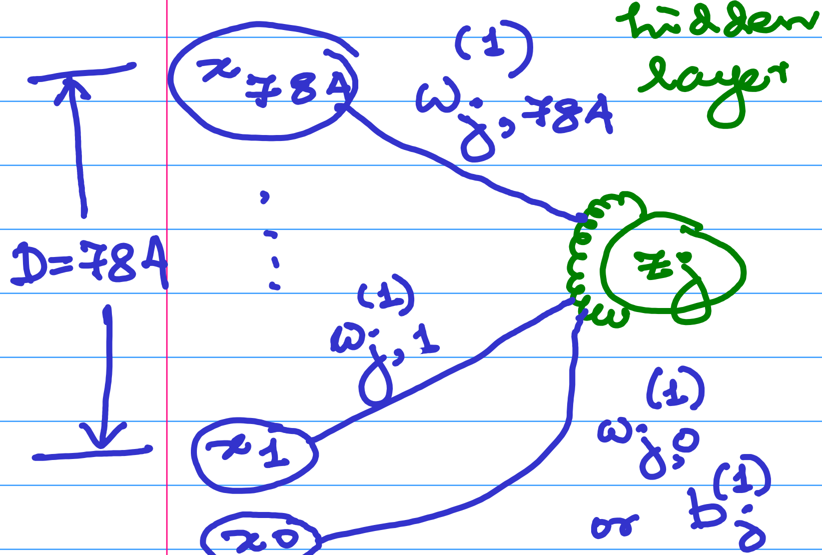
Basic structure: an MLP with 2 hidden layers

→ of specific interest is the first layer of connections

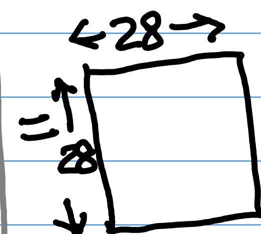


784×1 input vector

$w_{ji}^{(1)} \in \mathbb{R}$, a real number

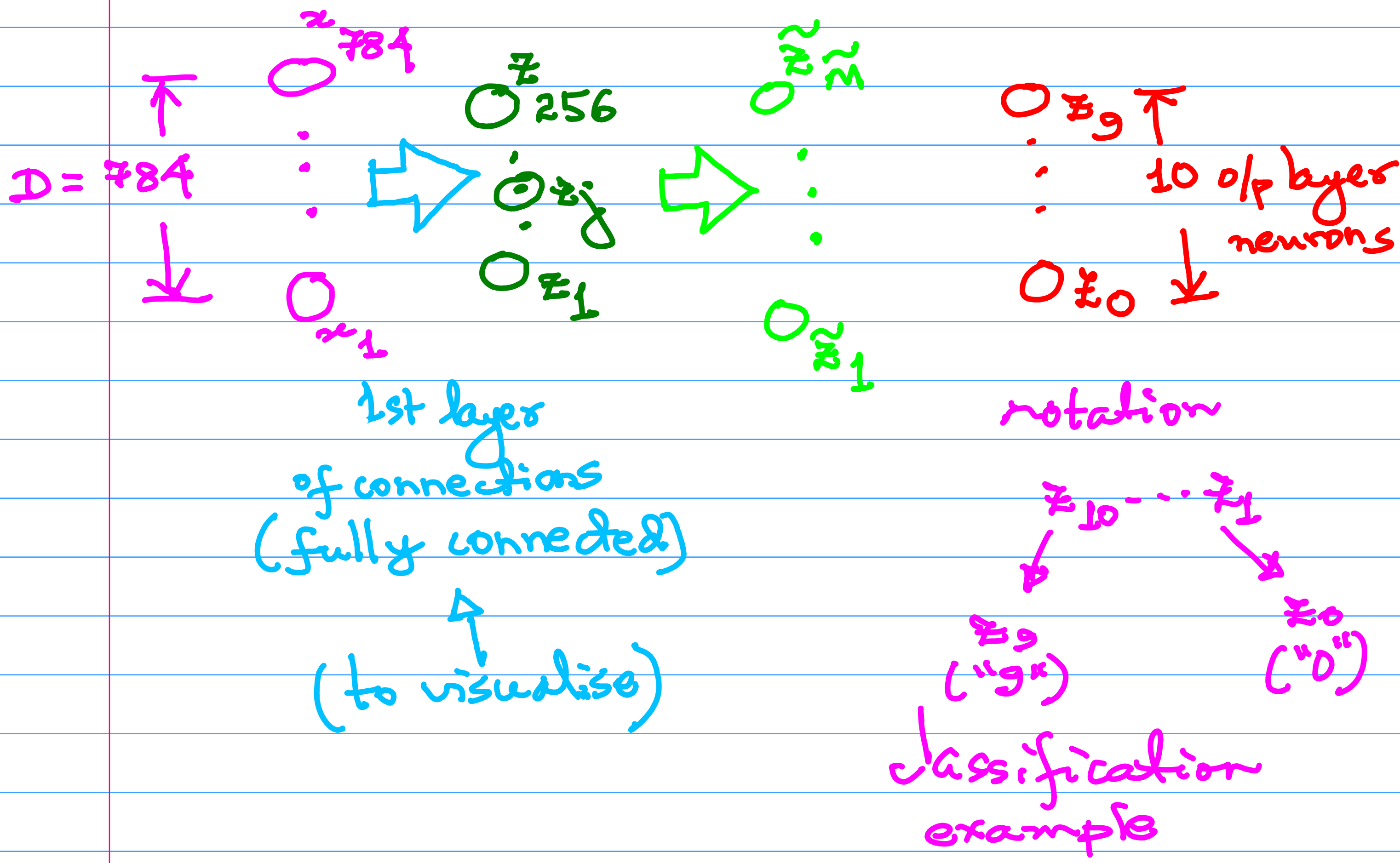


$$\underset{784\times 1}{\underline{w}^{(1)}} = \underset{784\times 1}{\begin{bmatrix} w_{j,784}^{(1)} \\ \vdots \\ w_{j,1}^{(1)} \end{bmatrix}}$$



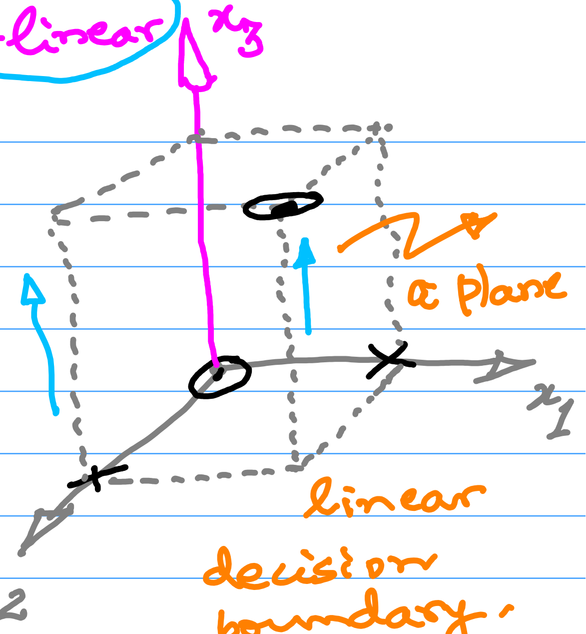
= "image" of the weights

Take-home point: The input is not an ordinary 1-D vector, but a 2-D image $\Rightarrow$ we can visualise the weights not as an ordinary 1-D vector, but a 2-D image with a similar spatial configuration/arrangement as the input itself.



original features		new feature	
x_2	x_1	$x_3 = x_2 x_1$	$y = x_2 \oplus x_1$
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

typically non-linear



The earlier (1, 1) point now "floats up" leading to an infinite number of planes (linear decision boundaries in 3-D) now separating the two classes (much like the concentric circles doughnut 'floating up' over the vada)

The 'Factorisation' in Math/Summation

Where does this appear, and why? Short Answer: Everywhere!

Working Rule:-
try putting it everywhere, and remove it if it is not required!

→ Probability (conditional/non-conditional)

→ Chain Rule (Calculus)
— total derivate/partial derivative.

PROBABILITY

The general probability factorisation

$$\begin{array}{l} \text{cause } 1 \searrow \\ \quad \vdots \qquad\quad \text{Effect} \\ \text{cause } N \nearrow \end{array}$$

Bayes Rule

$$p(\text{cause}\#i \mid \text{effect}) = \frac{p(\text{effect} \mid \text{cause}\#i)\, p(\text{cause}\#i)}{p(\text{effect})}$$

Symmetrical

$$\underbrace{p(A|B)\,p(B)}_{p(A \text{ and } B)} = \underbrace{p(B|A)\,p(A)}_{p(B \text{ and } A)}$$

$$\boxed{p(A|B)} = \frac{p(B|A)\,\boxed{p(A)}}{p(B)}$$

a priori probability of A

a posteriori probability of A

updated probability of A

$$x = x + 1$$

$$\Rightarrow x_{new} = x_{old} + 1$$

$$p(A)_{updated} = [\quad] \times p(A)_{initial}$$

$p(A)$

$p(A|B)$

→ As such, there is no difference between a conditional probability and an unconditional probability → these are just the updated and initial variants of the same physical quantity:

$$p(A|B) = \frac{p(B|A)\, p(A)}{p(B)}$$

$$\boxed{p(A|B,C)} = \frac{p(B|A,C)\,\boxed{p(A|C)}}{p(B|C)}$$

updated prob of A

initial prob of A

$\underset{\text{updated}}{p(A)} \leftarrow \boxed{\text{Bayes Rule}} \leftarrow \underset{\text{initial}}{p(A)}$

Where do the summations come in

$$p(\text{effect}) = \sum_{\forall j} p(\text{effect} \mid \text{cause}\#j)\, p(\text{cause}\#j)$$

→ safely put a summation for all j's

Example: →

tap on → lawn wet

rain → lawn wet

Suppose we find the lawn to be wet in the morning.

[Bayes Rule] $p\left(\begin{matrix}\text{rained}\\ \text{last night}\end{matrix} \middle| \begin{matrix}\text{lawn wet}\\ \text{in the morning}\end{matrix}\right)$

$$p(\text{rain} \mid \text{wet}) = \frac{p(\text{wet} \mid \text{rain})\, p(\text{rain})}{p(\text{wet})}$$

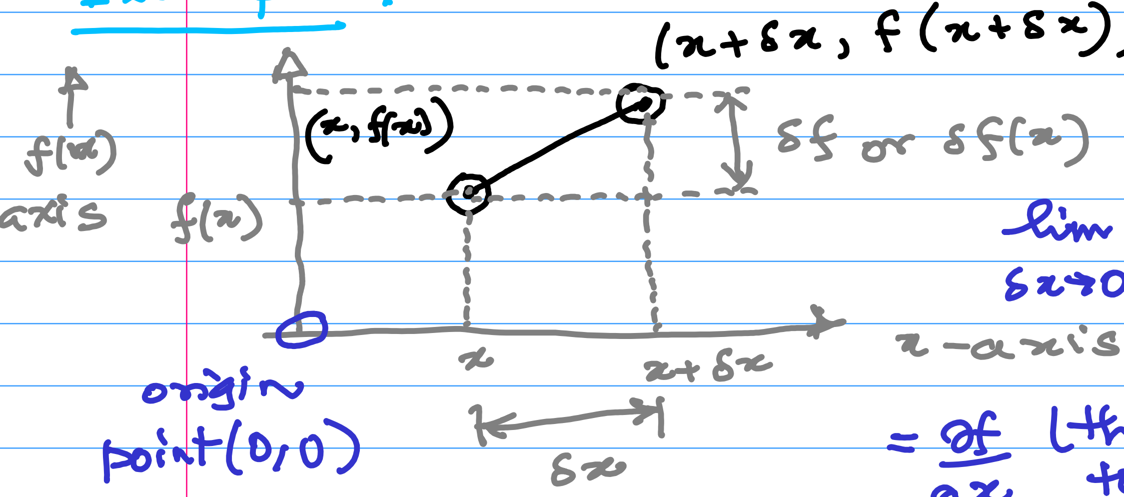
$$= \frac{p(\text{wet} \mid \text{rain})\, p(\text{rain})}{p(\text{wet} \mid \text{rain})\, p(\text{rain}) + p(\text{wet} \mid \text{tap on})\, p(\text{tap on})}$$

DERIVATIVES

The most general derivative is <u>NOT</u> the total derivative, but the partial derivatives.

$\boxed{\text{1-D}}$ one independent variable x → scalar
one dependent variable $f(x)$ → scalar function

Example: e.g., audio signal $f(t)$



$$\lim_{\delta x \to 0} \frac{f(x+\delta x) - f(x)}{(x+\delta x) - (x)}$$

$= \frac{\partial f}{\partial x}$ (this is also the total derivative df/dx)

Why? 1 scalar variable.

$$\lim_{\delta x \to 0} : \underbrace{f(x+\delta x) - f(x)}_{\delta f \text{ or } \delta f(x)} = \left(\frac{\partial f}{\partial x}\right) \underbrace{\delta x}$$

small change in the dependent variable, or the function

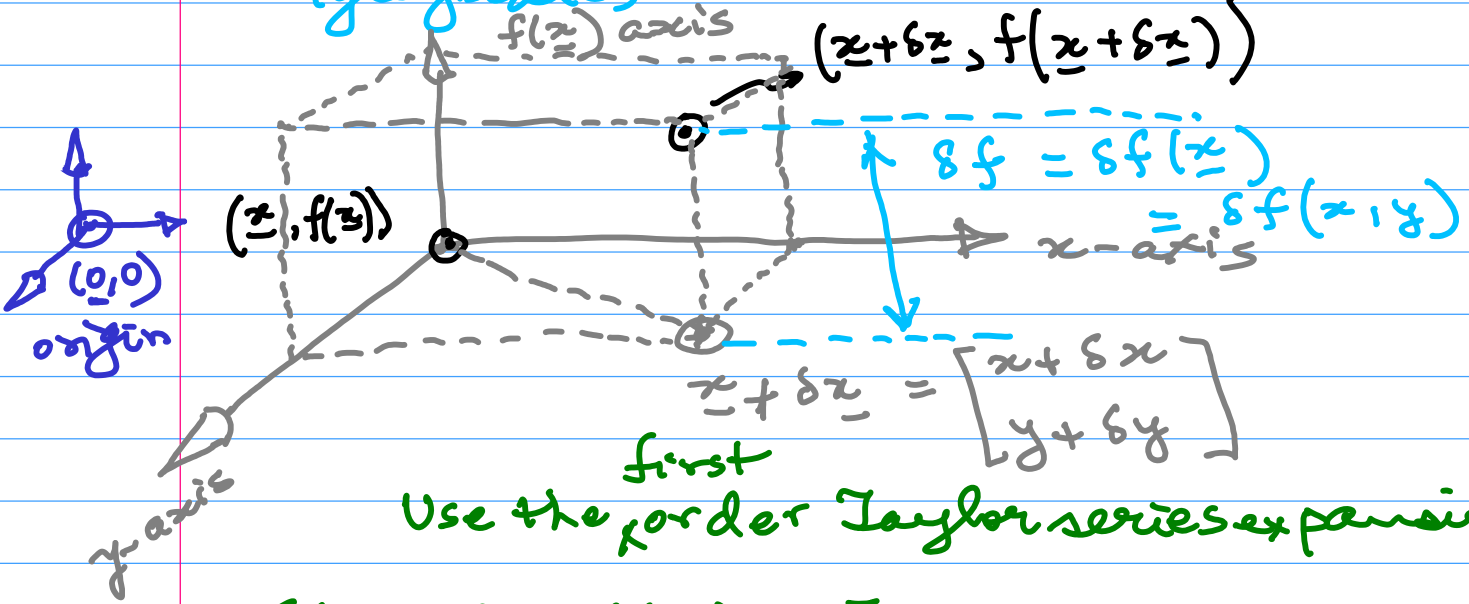
small change in the independent variable.

$\boxed{\text{2-D}}$ two independent variables $\underline{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ → vector

one dependent variable $\underbrace{f(\underline{x})}$ → scalar.

Example $I(x, y)$ or $I(\underline{x})$

the intensity of a pixel at position $\underline{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ (greyscale)



$$\underline{x}+\delta\underline{x} = \begin{bmatrix} x+\delta x \\ y+\delta y \end{bmatrix}$$

Use the first order Taylor series expansion

$$f(\underline{x}+\delta\underline{x}) = f(\underline{x}) + \bar{\nabla}_f \cdot \delta\underline{x}$$

$$\underbrace{f(\underline{x}+\delta\underline{x}) - f(\underline{x})}_{\delta f \text{ or } \delta f(\underline{x}) \text{ or } \delta f(x,y)} = \bar{\nabla} f \cdot \delta\underline{x}$$

$$\underbrace{\begin{bmatrix} \partial f/\partial x \\ \partial f/\partial y \end{bmatrix} \cdot \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}}_{\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y}$$

$$\underbrace{\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y}_{= \sum\limits_{var:\, x, y} \left(\frac{\partial f}{\partial\, var}\right)(\delta\, var)}$$

$\boxed{3\text{-D}}$ Three independent variables $\underline{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ → vector

one dependent variable $f(\underline{x})$ → scalar

Example: video (greyscale) $I(x, y, t)$: Intensity at pixel (x, y) in frame # t.