ITCS 5356: Mathematical Foundations

Derivatives

Razvan C. Bunescu

Department of Computer Science @ CCI

rbunescu@uncc.edu

Derivatives and Machine Learning

- For most ML algorithms, training means finding params $\mathbf{w}$ that minimize a *cost* function $J(\mathbf{w})$.
  - Example for linear regression:

$$J(\mathbf{w}) = \frac{1}{2N} \sum_{n=1}^{N} \boxed{(\mathbf{w}^T \mathbf{x}_\mathrm{n} - y_n)^2}$$

loss $l(\mathbf{x}_n; \mathbf{w})$

  - $l(\mathbf{x}_n; \mathbf{w})$ expresses the *loss* that a model with params $\mathbf{w}$ incurs on example $\mathbf{x}_n$.
    - When model classifies $\mathbf{x}_n$ correctly, loss will be defined to be low.
    - When model misclassifies $\mathbf{x}_n$, loss will be defined to be high.
  - We will see in this course how to formulate $l(\mathbf{x}_n; \mathbf{w})$ to make the cost $J(\mathbf{w})$ easy to minimize on a computer.

Derivatives and Machine Learning

- For most ML algorithms, training means finding params $\mathbf{w}$ that minimize a *cost* function $J(\mathbf{w})$.
  - Example for linear regression:

$$J(\mathbf{w}) = \frac{1}{2N}\sum_{n=1}^{N} \boxed{(\mathbf{w}^T\mathbf{x}_\mathrm{n} - y_n)^2}$$

loss $l(\mathbf{x}_n; \mathbf{w})$

  - If the *cost* $J(\mathbf{w})$ is differentiable, at the minimum its derivative is 0.
    - We use numerical algorithms (SGD) to find $\mathbf{w}$ such that the derivative of the loss $\frac{\partial J}{\partial \mathbf{w}} = 0$

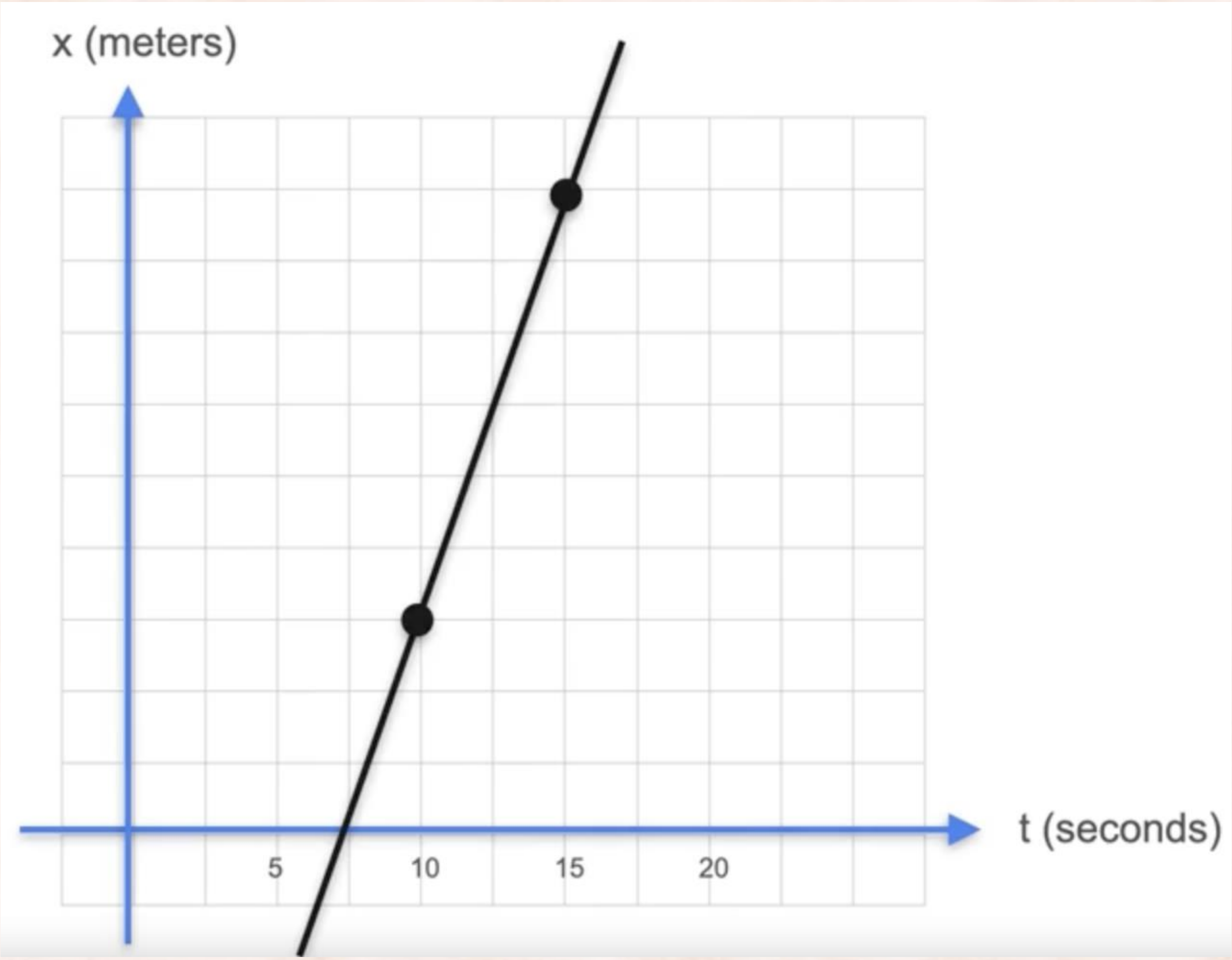
=> <u>we need to understand derivatives</u>.

Univariate Functions

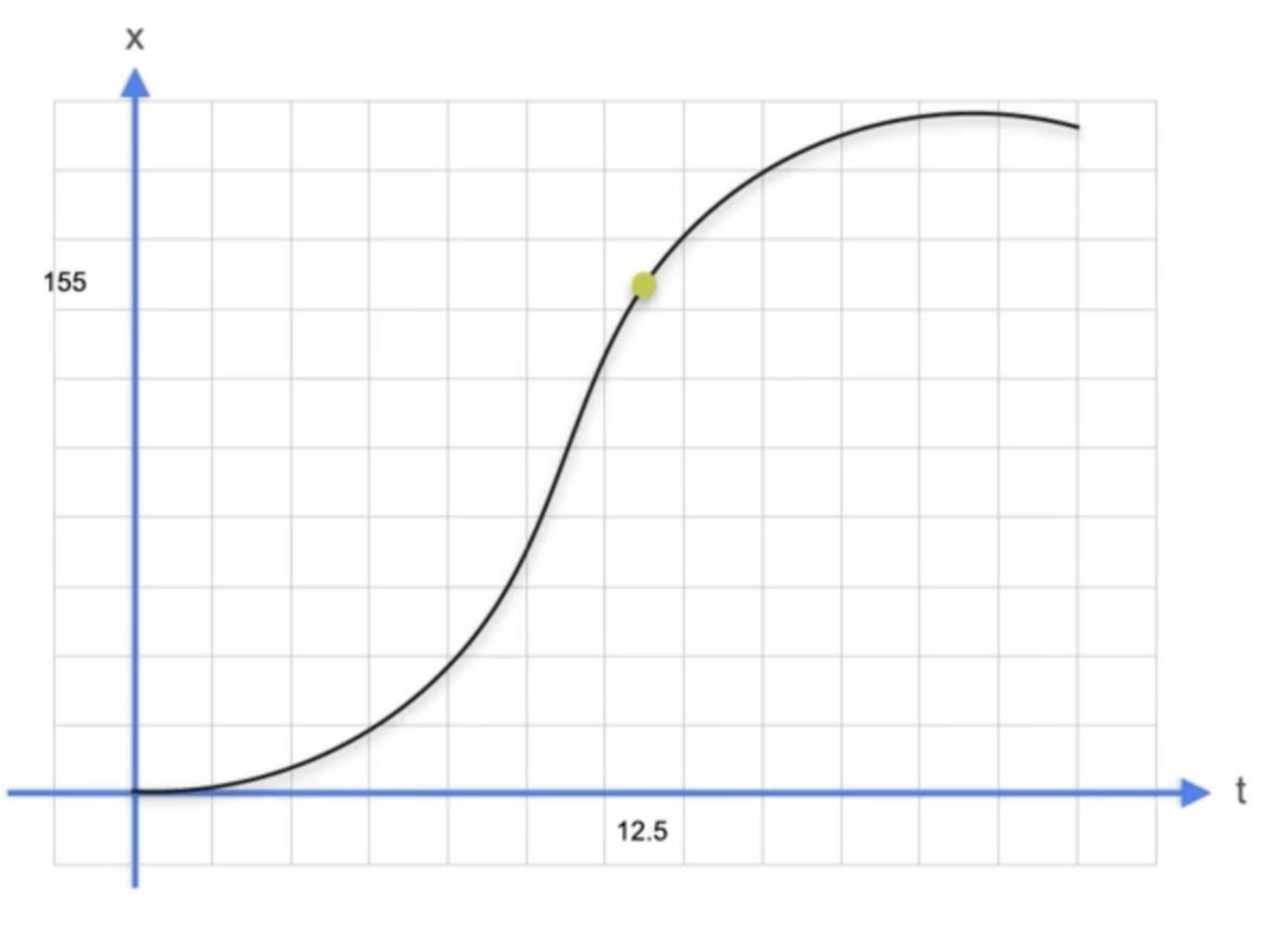
- A univariate function means has one scalar as input:
  - We can represent univariate functions on a 2D graph, with input on the horizontal axis and output (function value) on the vertical axis.

- For example:
  - Position x is a function of time t, so we can write is as $x(t)$.
  - Pressure p is a function of temperature t, hence $p(t)$.

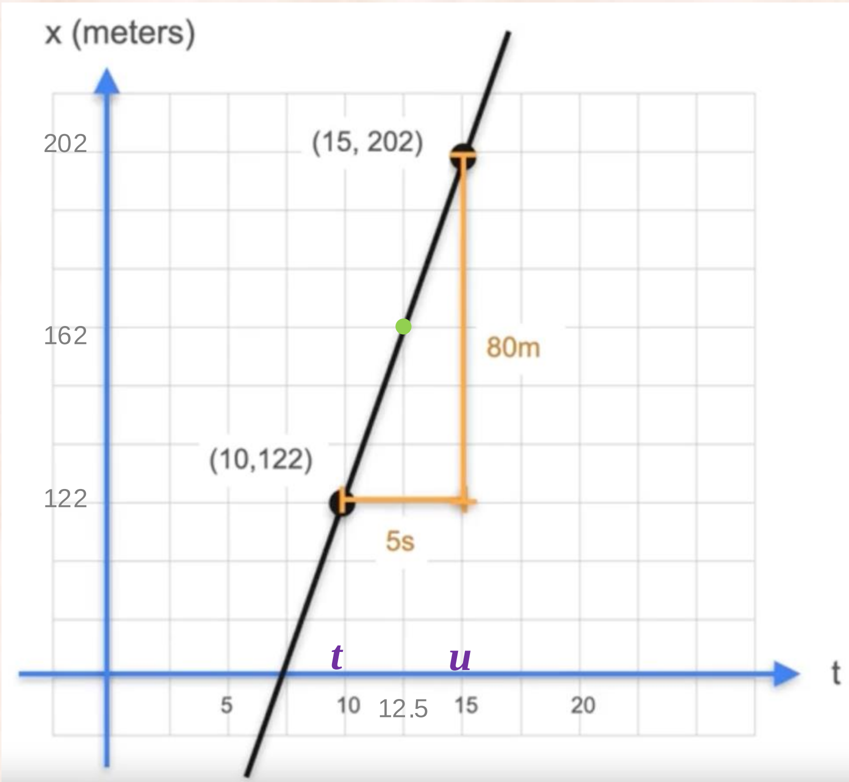
$x(t)$: position x as a function of time t



$x(t)$: position x as a function of time t



$x'(t)$: speed is derivative of x with respect to t



- Speed is the *rate of change* of x with respect to t.

$$x'(t) = \frac{\Delta x}{\Delta t} = \frac{x(u) - x(t)}{u - t}$$

$$x'(10) = \frac{x(15) - x(10)}{15 - 10} = \frac{80}{5} = 16$$

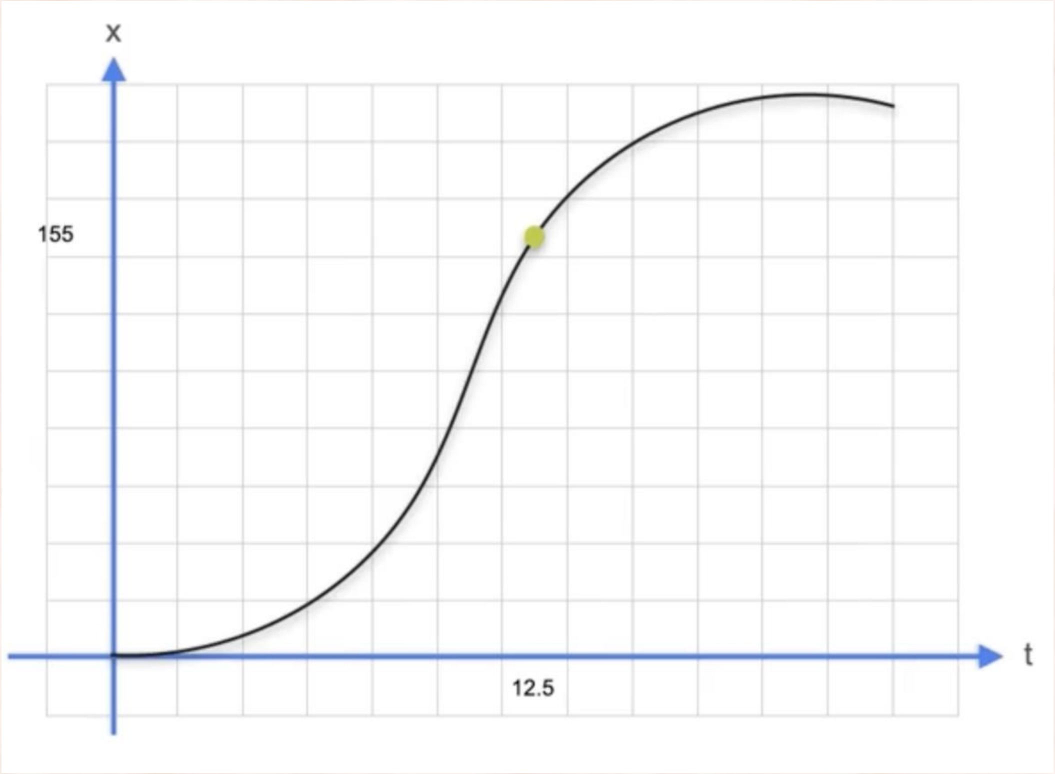
$$x'(10) = \frac{x(12.5) - x(10)}{12.5 - 10} = \frac{40}{2.5} = 16$$

$x'(12.5) = ?$ $\quad$ $x'(15.0) = ?$

$x'(1024) = ?$ $\quad$ $x'(1024) = ?$

- Speed is the *slope* of the line $x(t)$ at time t.
  - What if the graph is not a line? *Hint: trigonometric name for the slope …*

$x'(t)$: speed is derivative of x with respect to t

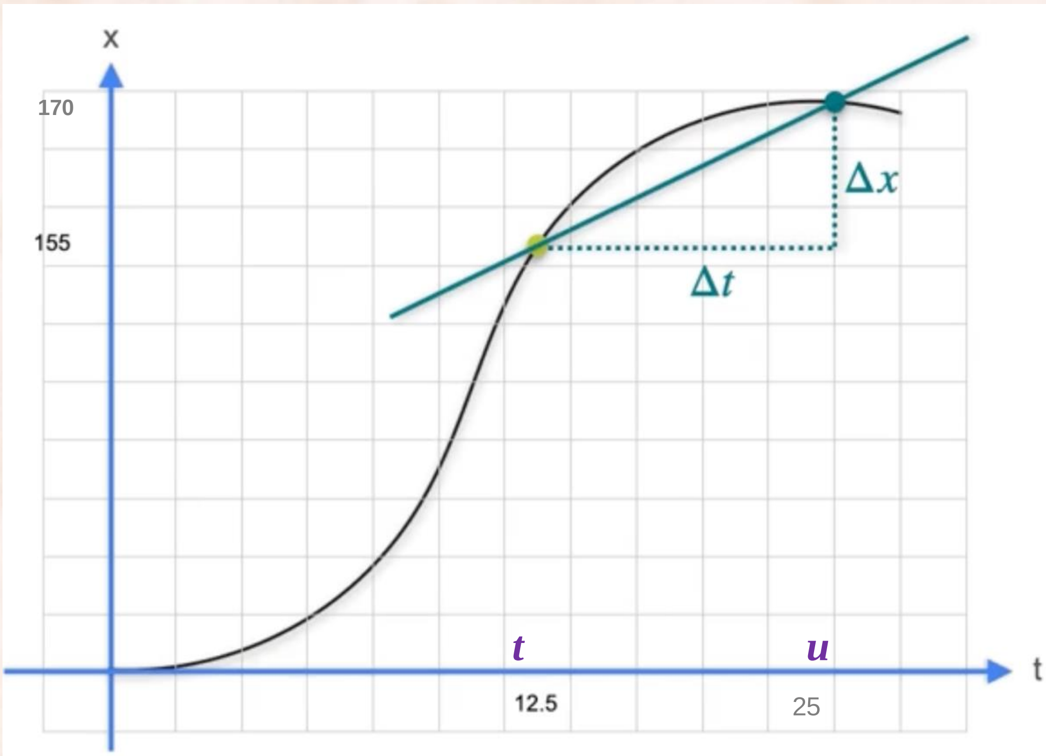


$$x'(t) = \frac{\Delta x}{\Delta t} = \frac{x(u) - x(t)}{u - t}$$

$t = 12.5$

$u = ?$

$x'(t)$: speed is derivative of x with respect to t



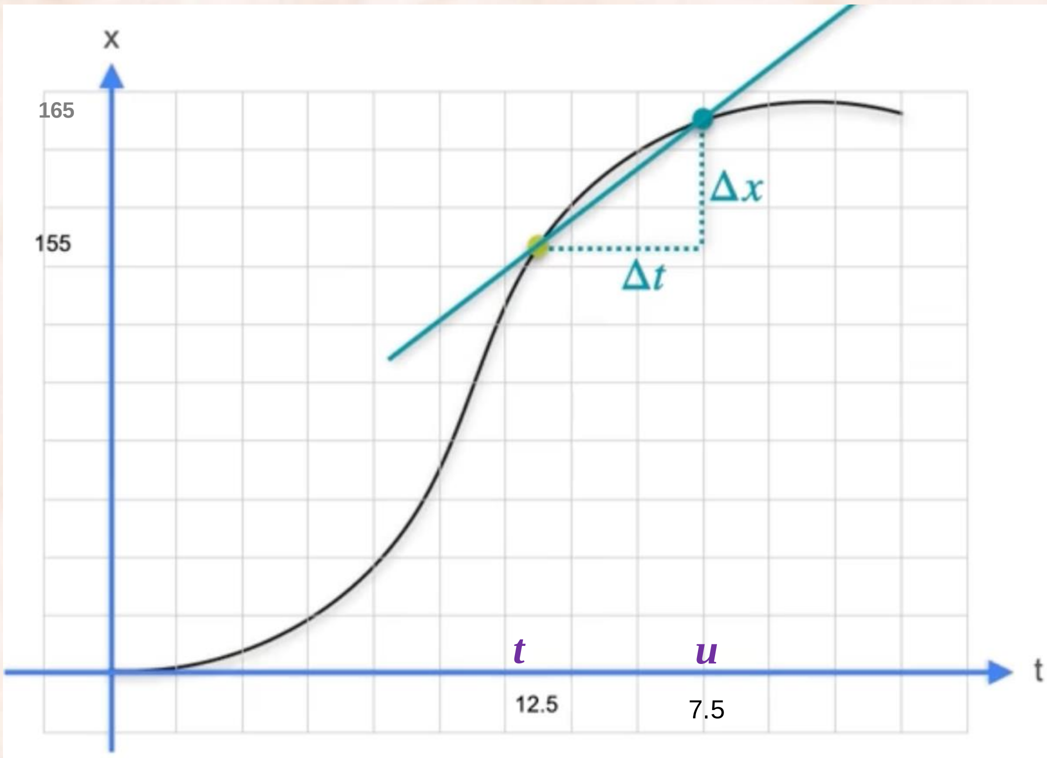
$$x'(t) = \frac{\Delta x}{\Delta t} = \frac{x(u) - x(t)}{u - t}$$

$$t = 12.5$$

$$u = t + \mathbf{12.5} = 25.0$$

$$x'(t) = \frac{15.0}{12.5} = 1.20$$

$x'(t)$: speed is derivative of x with respect to t



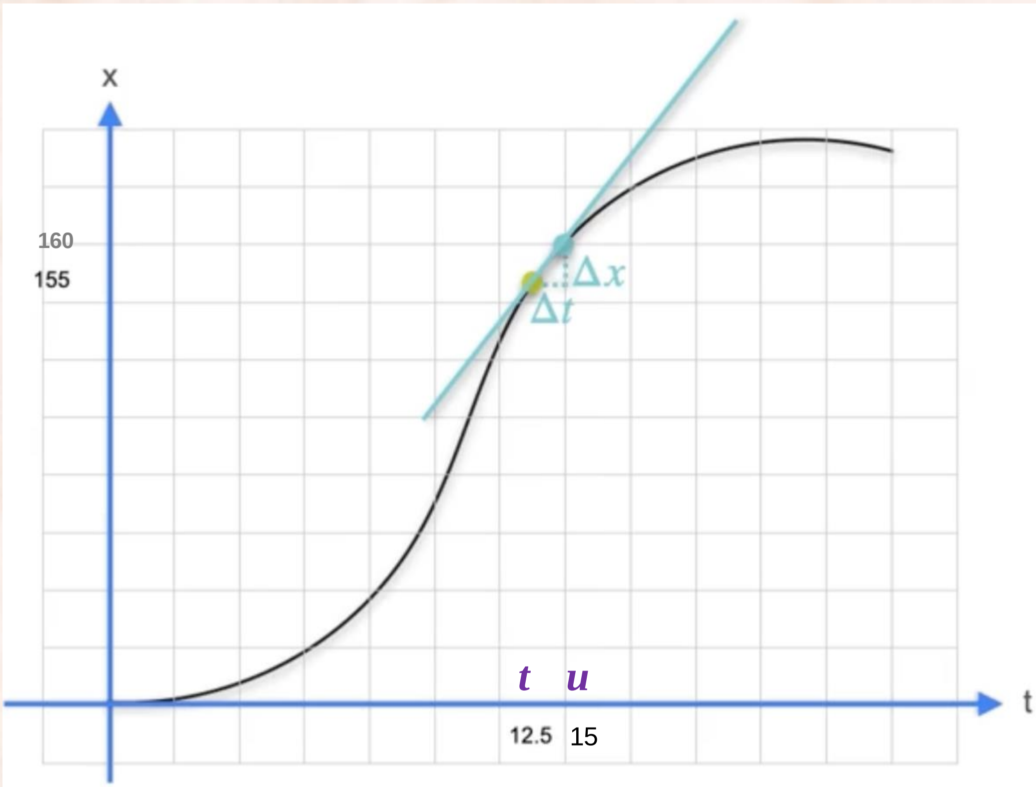
$$x'(t) = \frac{\Delta x}{\Delta t} = \frac{x(u) - x(t)}{u - t}$$

$$t = 12.5$$

$$u = t + \mathbf{7.5} = 20.0$$

$$x'(t) = \frac{10.0}{7.5} = 1.33$$

$x'(t)$: speed is derivative of x with respect to t



$$x'(t) = \frac{\Delta x}{\Delta t} = \frac{x(u) - x(t)}{u - t}$$

$$t = 12.5$$

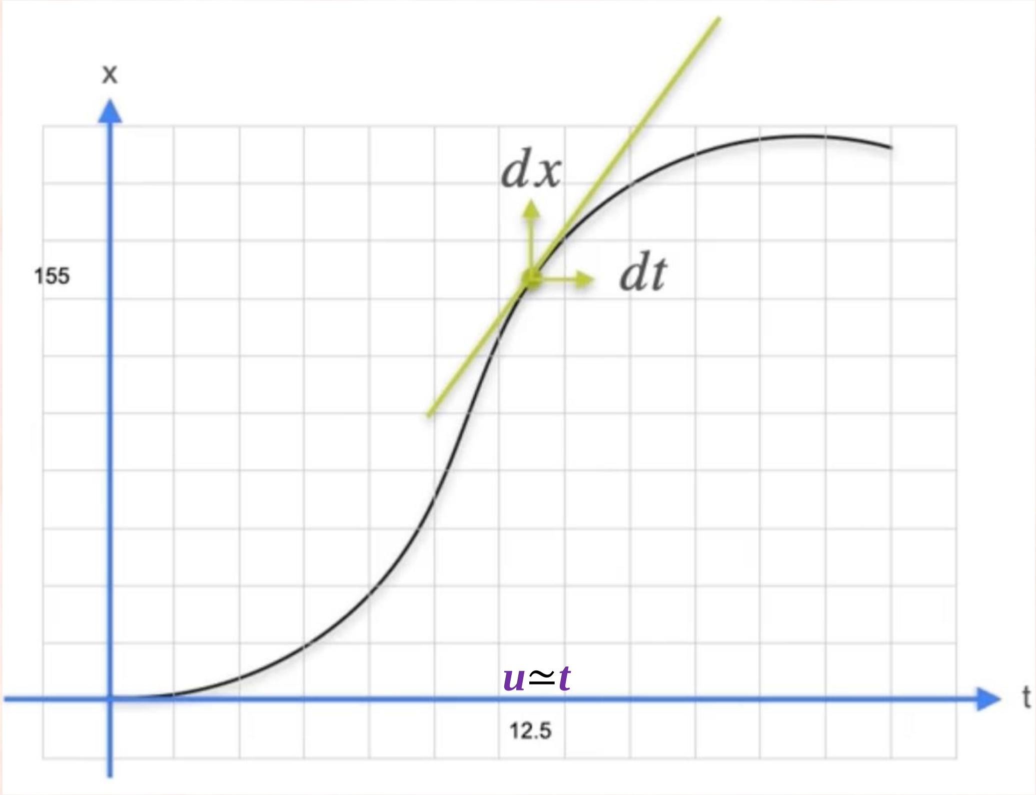
$$u = t + \mathbf{2.5} = 15.0$$

$$x'(t) = \frac{5.0}{2.5} = 2.0$$

Let's make u arbitrarily close to t:

$$u = t + \boldsymbol{\varepsilon} \simeq 12.5$$

$x'(t)$: speed is derivative of x with respect to t



$$x'(t) = \frac{\Delta x}{\Delta t} = \frac{x(u) - x(t)}{u - t}$$

where $u = t + \boldsymbol{\varepsilon}$

$$x'(t) = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{x(t+\varepsilon) - x(t)}{(t+\epsilon) - t}$$

$$= \frac{x(t+\varepsilon) - x(t)}{\epsilon}$$

where ε *is very close to* 0

Derivative $x'(t) = \lim_{\epsilon \to 0} \frac{x(t+\varepsilon) - x(t)}{\varepsilon}$

Formal definition of derivative

- Given a univariate function $f(x)$, the **derivative** of f with respect to x, when evaluated at x_0, is defined as:

$$f'(x_0) = \lim_{\epsilon \to 0} \frac{f(x_0 + \varepsilon) - f(x_0)}{\varepsilon}$$

- Observations:
  - When this limit does not exist => f is not differentiable at x_0.
    - Can you think of an example?
  - How do we calculate it?
    - Use the definition, L'Hopital, differentiation rules, …

Examples: Constant functions

- Compute the derivative of $f(x) = c$, where c is a constant.

$$f'(x) = \lim_{\epsilon \to 0} \frac{f(x + \varepsilon) - f(x)}{\varepsilon}$$

$$= \lim_{\epsilon \to 0} \frac{c - c}{\varepsilon} = \lim_{\epsilon \to 0} 0 = 0$$

Examples: Linear functions

- Compute the derivative of $\boldsymbol{f(x) = ax + b}$, where a and b are constants.

$$\boldsymbol{f'(x)} = \lim_{\epsilon\to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

$$= \lim_{\epsilon\to 0} \frac{a(x+\varepsilon) + b - (ax+b)}{\varepsilon}$$

$$= \lim_{\epsilon\to 0} \frac{ax + a\varepsilon + b - ax - b}{\varepsilon}$$

$$= \lim_{\epsilon\to 0} \frac{a\varepsilon}{\varepsilon} = \lim_{\epsilon\to 0} a = \boldsymbol{a}$$

Examples: Quadratic terms

- Compute the derivative of $\boldsymbol{f(x) = ax^2}$, where *a* and *b* are constants.

$$\boldsymbol{f'(x)} = \lim_{\epsilon\to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

$$= \lim_{\epsilon\to 0} \frac{a(x+\varepsilon)^2 - ax^2}{\varepsilon}$$

$$= \lim_{\epsilon\to 0} \frac{ax^2 + 2a\varepsilon x + a\varepsilon^2 - ax^2}{\varepsilon}$$

$$= \lim_{\epsilon\to 0} 2ax + a\varepsilon = 2ax + a\lim_{\epsilon\to 0} \varepsilon = 2ax + 0 = \boldsymbol{2ax}$$

Notation for Derivatives

Who	First	Second
Leibniz	$\frac{\mathbf{d}\boldsymbol{f}}{\mathbf{d}\boldsymbol{x}}$ or $\frac{\boldsymbol{\delta f}}{\boldsymbol{\delta x}}$	$\frac{d^2 f}{dx^2}$ or $\frac{\delta^2 f}{\delta x^2}$
Lagrange	$\boldsymbol{f'(x)}$	$f''(x)$
Euler	$D_x f$	$D_x^2 f$
Newton	$\dot{f}$	$\ddot{f}$

- For **univariate** functions $f(x)$, we'll use Leibniz and Lagrange.
- For **multivariate** functions $f(x_1, x_2, \ldots, x_n)$ we'll use Lagrange:
  - $\frac{\delta f}{\delta x_k}$ is the **partial derivative** of f with respect to x_k.
  - The vector $\nabla f(\mathbf{x}) = \frac{\delta f}{\delta \mathbf{x}} = \left[\frac{\delta f}{\delta x_1}, \frac{\delta f}{\delta x_2}, \ldots, \frac{\delta f}{\delta x_n}\right]$ is called the **gradient** of f with respect to the vector of parameters $\mathbf{x} = [x_1, x_2, \ldots, x_n]$.

Properties of Derivatives

Using the definition, prove that:

- $\boldsymbol{h(x) = cf(x)}$, then $\boldsymbol{h'(x) = cf'(x)}$.
- $\boldsymbol{h(x) = f(x) + g(x)}$, then $\boldsymbol{h'(x) = f'(x) + g'(x)}$.

Using the above, prove that if $h(x) = af(x) + b$, then $h'(x) = af'(x)$.

Derivatives of multiplications and divisions of univariate functions:

- $h(x) = f(x)g(x)$, then $h'(x) = f'(x)g(x) + f(x)g'(x)$
- $h(x) = \frac{f(x)}{g(x)}$, then $h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$

Using the above, prove that if $h(x) = \frac{1}{g(x)}$, then $h'(x) = -\frac{1}{g^2(x)}$

Derivatives of Common Functions

$h(x)$	$h'(x)$
a	0
x^a	ax^{a-1}
e^x	e^x
$\ln(x)$	$1/x$
$sin(x)$	$cos(x)$
$cos(x)$	$-sin(x)$

Univariate Chain Rule for Differentiation

- Univariate Chain Rule:

$$f = f \circ g \circ h = f(g(h(x)))$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g} \frac{\partial g}{\partial h} \frac{\partial h}{\partial x}$$

- Example:

$$f(g(x)) = 2g(x)^2 - 3g(x) + 1$$

$$g(x) = x^3 + 2x$$

Multivariate Chain Rule for Differentiation

- Multivariate Chain Rule:

$$f = f(g_1(x), g_2(x), \ldots, g_n(x))$$

$$\frac{\partial f}{\partial x} = \sum_{i=1}^{n} \frac{\partial f}{\partial g_i} \frac{\partial g_i}{\partial x}$$

- Example:

$$f(g_1(x), g_2(x)) = 2g_1(x)^2 - 3g_1(x)g_2(x) + 1$$

$$g_1(x) = 3x$$

$$g_2(x) = x^2 + 2x$$

Exercises for Differentiation Rules

Exercises for Differentiation Rules