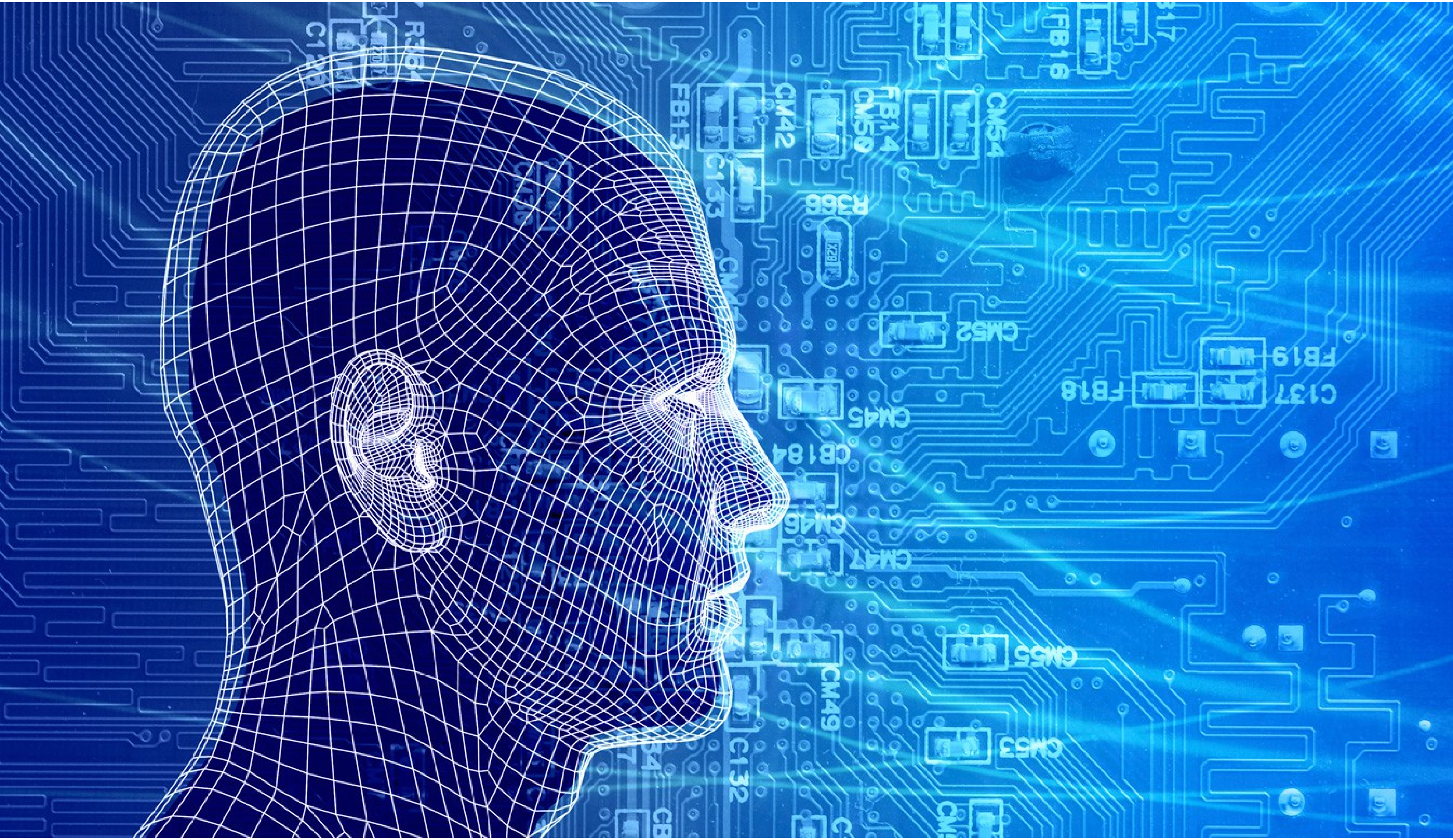
Can machines think like human beings and beyond!



1. If y = f(x), then

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

2. If u = f(x), v = f(x), then

- $\frac{d}{dx}(k) = 0$
- $\frac{d}{dx}(ku) = k\frac{du}{dx}$
- $\frac{d}{dx}(u \pm v) = \frac{du}{dx} \pm \frac{dv}{dx}$
- $\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$
- $\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} + u\frac{du}{dx}}{v^2}$
- If x = f(t), y = ϕ(t), then $\frac{dy}{dx} = \frac{dy}{dt} \Big/ \frac{dx}{dt}$
- If y =f[ϕ(x)], then $\frac{dy}{dx}$ =f'[f(x)]. $\frac{d}{dx}\left[\phi(x)\right]$
- $\frac{dy}{dx} . \frac{dx}{dy} = 1 \text{ or } \frac{dy}{dx} = \frac{1}{dx/dy}$

Differential calculus: power rule

3.1.8 The Power Rule

$$\frac{d}{dx}u^n = nu^{n-1}\frac{du}{dx}$$ for any power n, integer, rational or irrational.

hence,

$$\frac{d}{dx}x^n = nx^{n-1}$$

implies

$$n = 0: \frac{d}{dx}1 = 0$$

$$n = 1 : \frac{d}{dx}x = 1$$

$$n = 2: \frac{d}{dx}x^2 = 2x^{2-1} = 2x$$

Proof of power rule

Example of power rule

Gradient descent with sigmoid on a perceptron

First, notice $g'(x) = g(x)(1-g(x))$

Because: $g(x) = \frac{1}{1+e^{-x}}$ so $g'(x) = \frac{-e^{-x}}{\left(1+e^{-x}\right)^2}$

$$= \frac{1-1-e^{-x}}{\left(1+e^{-x}\right)^2} = \frac{1}{\left(1+e^{-x}\right)^2} - \frac{1}{1+e^{-x}} = \frac{-1}{1+e^{-x}}\left(1-\frac{1}{1+e^{-x}}\right) = -g(x)(1-g(x))$$

$$\text{Out(x)} = g\left(\sum_k w_k x_k\right)$$

$$\text{E} = \sum_i \left(y_i - g\left(\sum_k w_k x_{ik}\right)\right)^2$$

$$\frac{\partial \text{E}}{\partial w_j} = \sum_i 2\left(y_i - g\left(\sum_k w_k x_{ik}\right)\right)\left(-\frac{\partial}{\partial w_j} g\left(\sum_k w_k x_{ik}\right)\right)$$

$$= \sum_i -2\left(y_i - g\left(\sum_k w_k x_{ik}\right)\right) g'\left(\sum_k w_k x_{ik}\right) \frac{\partial}{\partial w_j} \sum_k w_k x_{ik}$$

$$= \sum_i -2\delta_i g(\text{net}_i)(1-g(\text{net}_i))x_{ij}$$

where $\delta_i = y_i - \text{Out}(\text{x}_i)$ $\quad \text{net}_i = \sum_k w_k x_k$

The sigmoid perceptron update rule:

$$w_j \leftarrow w_j + \eta \sum_{i=1}^{R} \delta_i g_i (1-g_i) x_{ij}$$

where $$g_i = g\left(\sum_{j=1}^{m} w_j x_{ij}\right)$$

$$\delta_i = y_i - g_i$$

Probability theory discussion

- Toss of a coin
- Deck of cards
- Draw of two
- Probability provided a condition
- Interesting anecdote