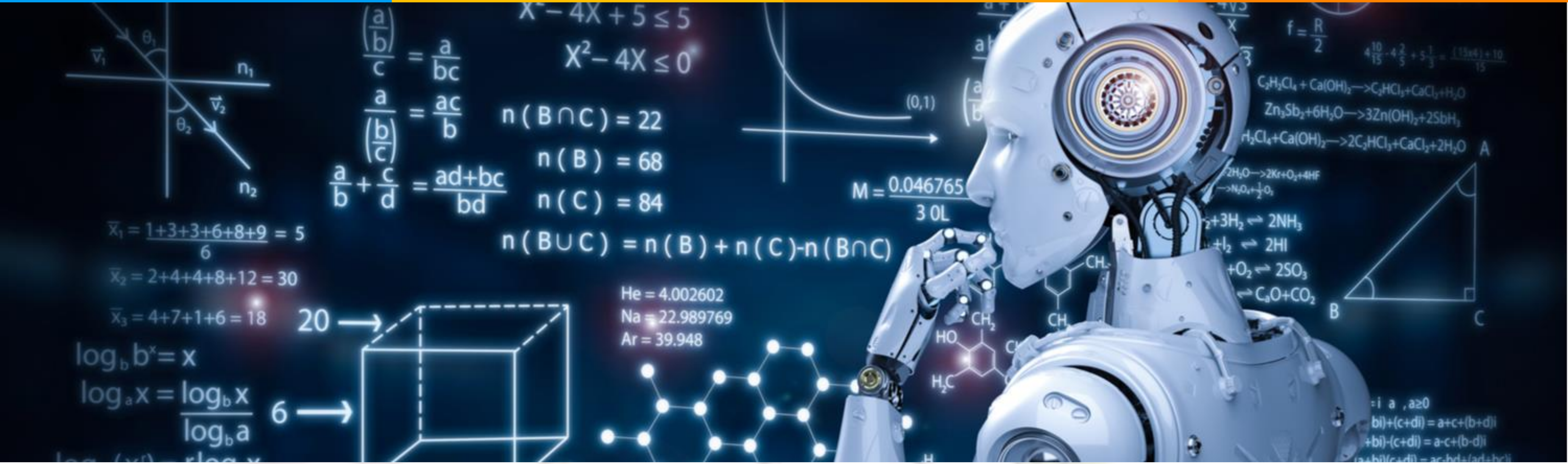
Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML



Lecture 10 Taylor’s approximation

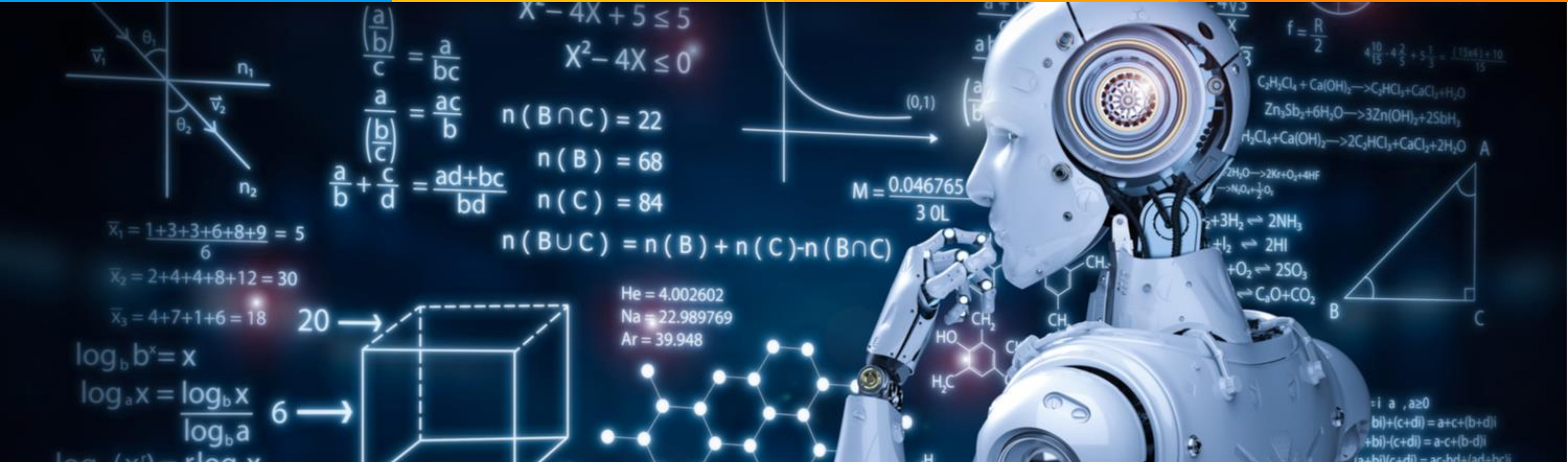
Recap

- Covered Gram Schmidt
- Finished Chapter 5

Why study Taylor’s approximation?

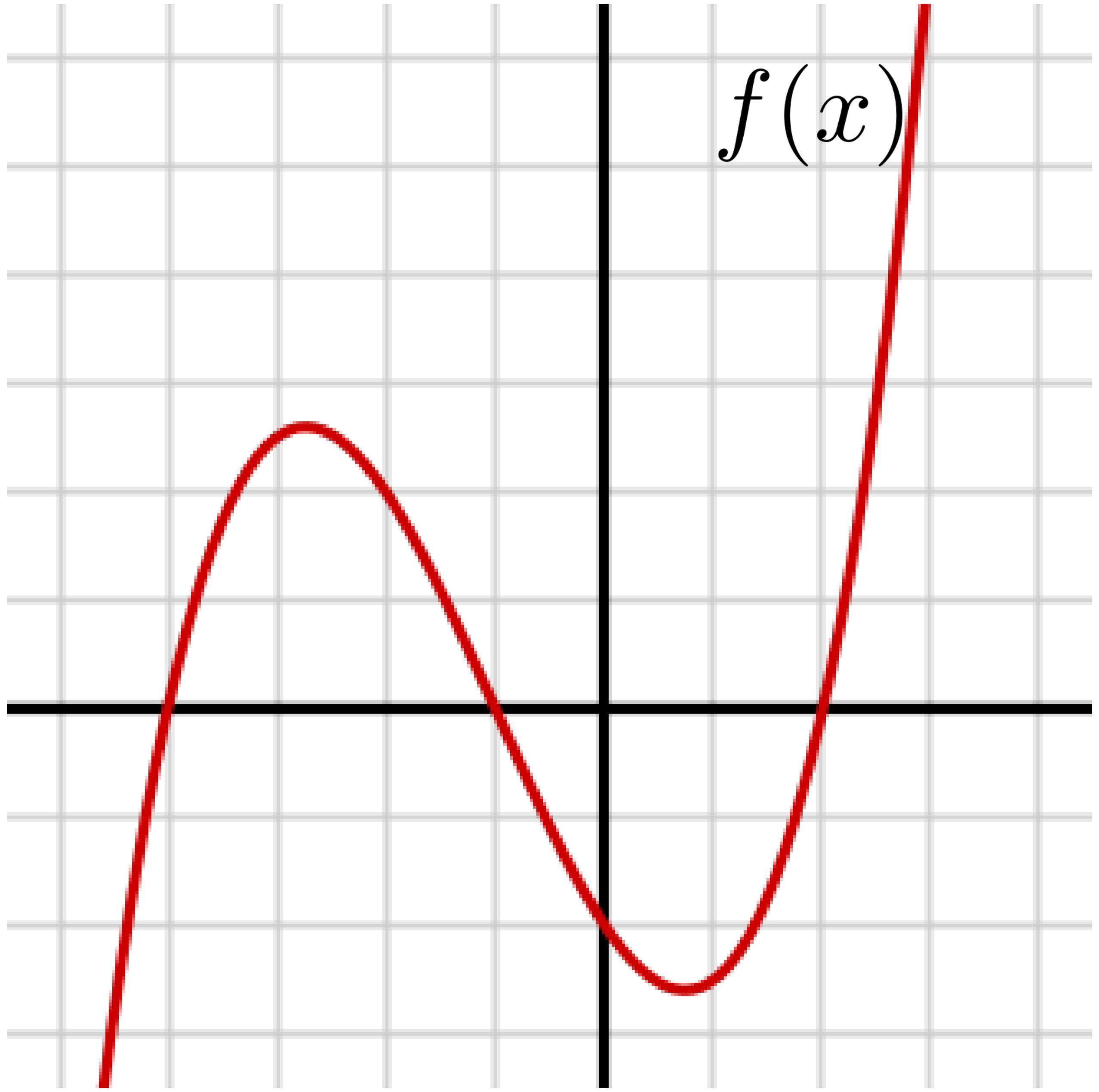
- Numerical method for
  - Gradient descent in Machine Learning is based on Taylor’s first order approximation
  - Newton Raphson’s method in convex optimization is based on Taylor’s second order approximation
- Single & multivariable intuition

Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML



Taylor's approximation & gradient descent in one variable

Function approximation at 0



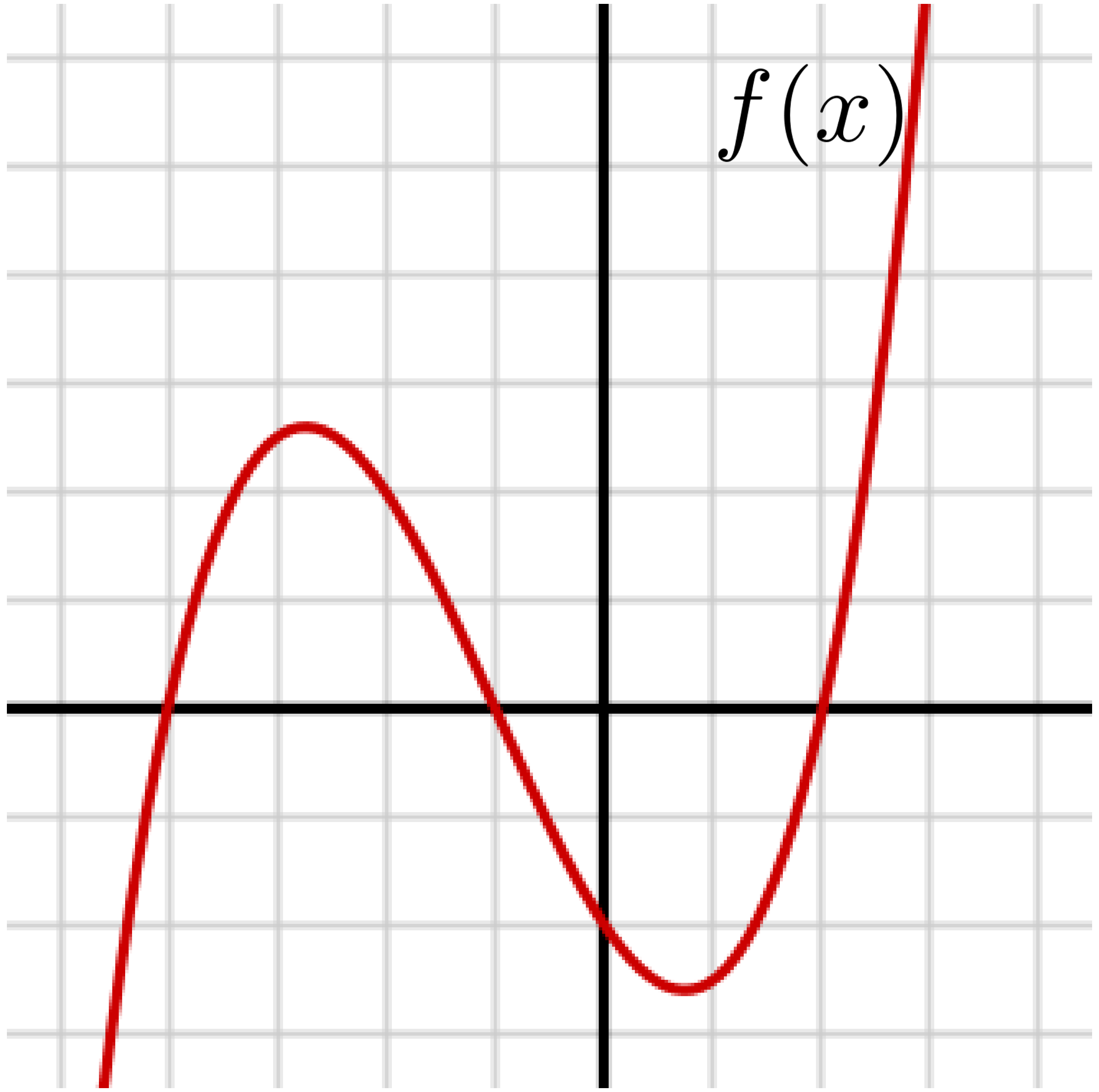
- f(x) is continuous, differentiable at x=0
- Approximate with a polynomial

$$p(0) = f(0) \qquad p(x) \approx f(0)$$

$$p'(0) = f'(0) \qquad p(x) \approx f(0) + f'(0)x$$

$$p''(0) = f''(0) \qquad p(x) \approx f(0) + f'(0)x + \frac{1}{2}f''(0)x^2$$

Function approximation at any constant z



- f(x) is continuous, differentiable at x=z
- Approximate with a polynomial

$$p(z) = f(z) \qquad\qquad p(x) \approx f(z)$$

$$p'(z) = f'(z) \qquad\qquad p(x) \approx f(z) + f'(z)(x - z)$$

$$p''(z) = f''(z) \qquad\qquad p(x) \approx f(z) + f'(z)x + \frac{1}{2}f''(z)(x - z)^2$$

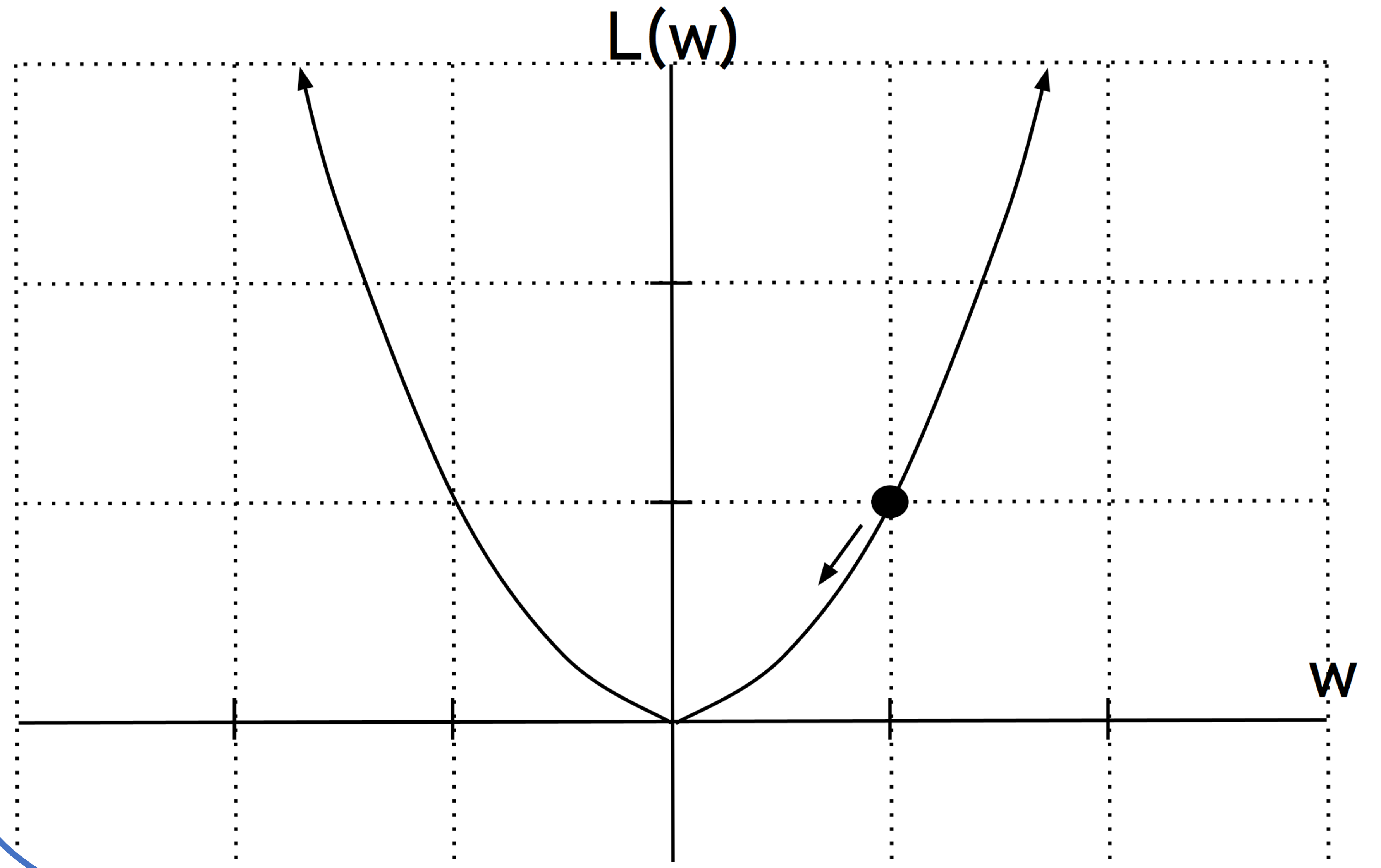
Univariate gradient descent

$$p(x) \approx f(z) + f'(z)(x - z)$$

$$\mathcal{L}(w) \approx \mathcal{L}(z) + \mathcal{L}'(z)(w - z)$$

$$w \to z; \qquad w - z = s$$

$$\mathcal{L}(\underbrace{z + s}_{w_{new}}) \approx \mathcal{L}(\underbrace{\;z\;}_{w_{old}}) + \mathcal{L}'(\underbrace{\;z\;}_{w_{old}})s$$



$$s = -\alpha \mathcal{L}'(w)$$

$$w_{new} = w_{old} - \alpha \mathcal{L}'(w_{old})$$

$$\mathcal{L}(w_{new}) \approx \mathcal{L}(w_{old}) - \alpha \mathcal{L}'(w_{old})^2 < \mathcal{L}(w_{old})$$

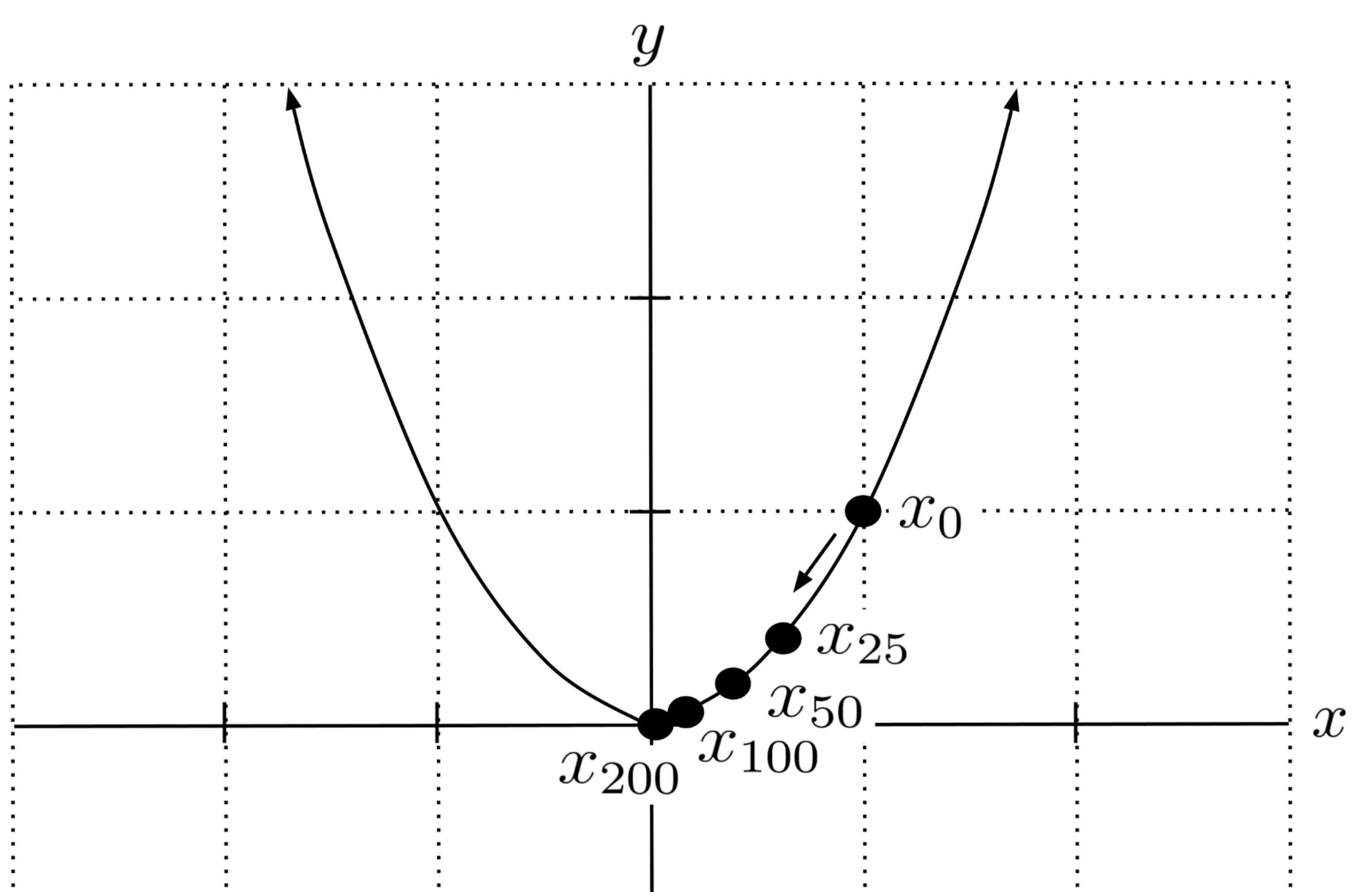
Less than holds only for convex function

Univariate gradient descent example

$$w_{new} = w_{old} - \alpha \mathcal{L}'(w_{old})$$

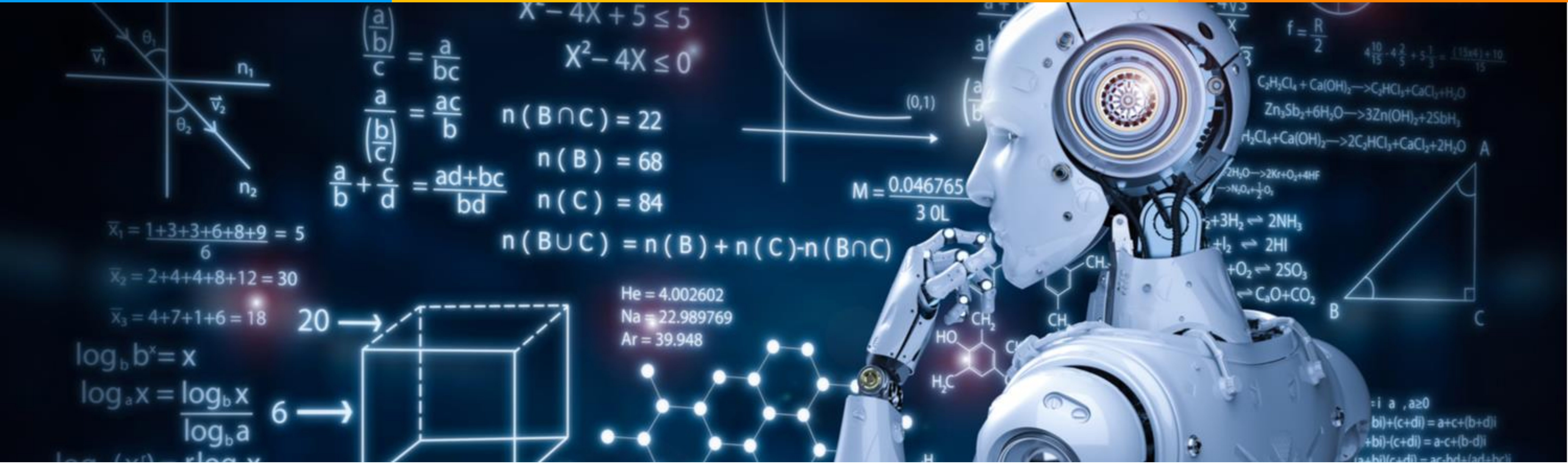
$$\alpha = 0.01$$

$$\mathcal{L}(w_{new}) \approx \mathcal{L}(w_{old}) - \alpha \mathcal{L}'(w_{old})^2$$



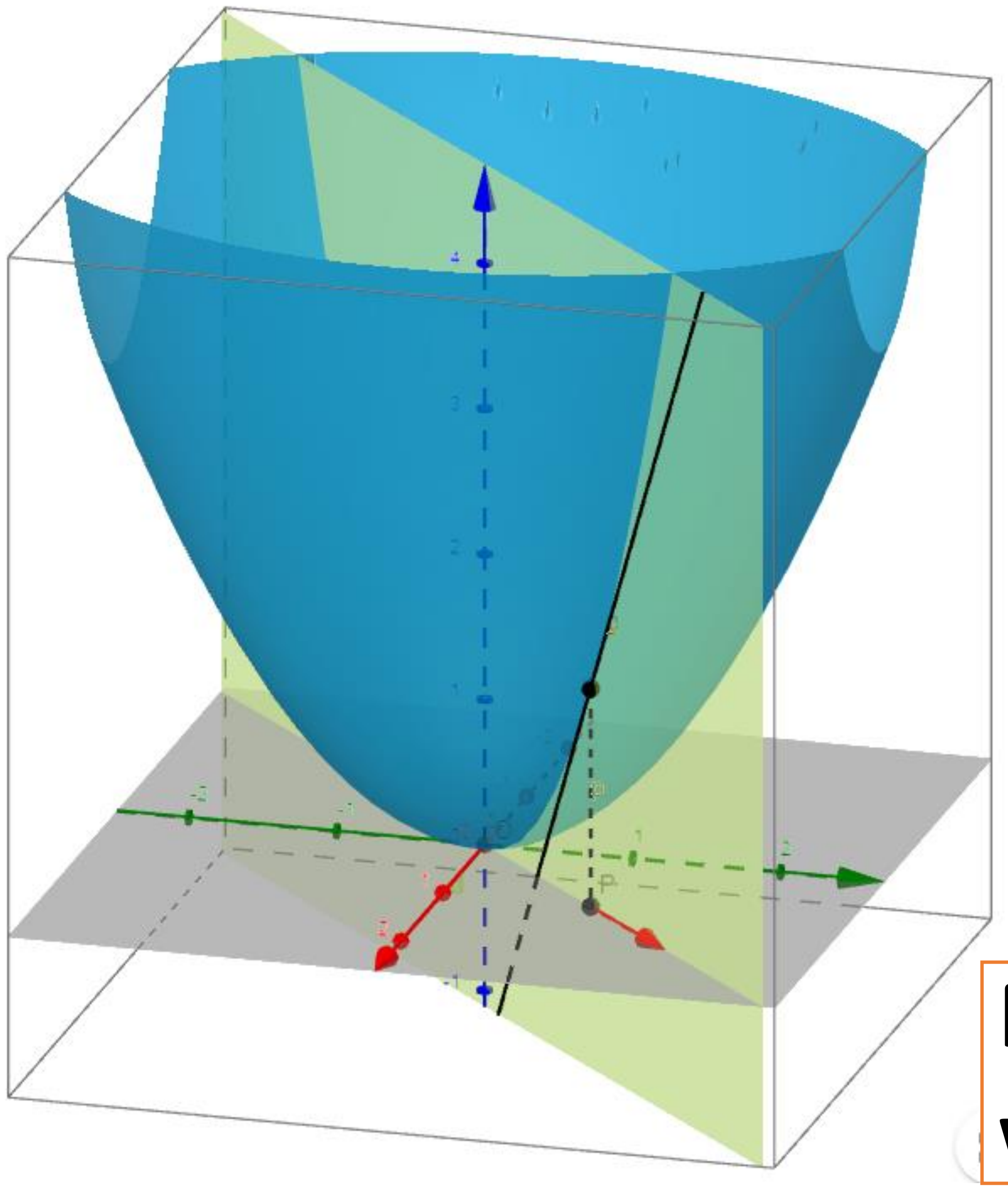
n	w_n	$\mathcal{L}'(w_n)$	$\alpha\mathcal{L}'(w_n)$
0	1	2	0.02
1	0.98	1.96	0.0196
2	0.9604	1.9208	0.019208
3	0.941192	1.882384	0.018824
⋮	⋮	⋮	⋮
25	0.603465	1.206929	0.012069
50	0.364170	0.728339	0.007283
100	0.132620	0.265239	0.002652
200	0.017588	0.035176	0.000352
300	0.002333	0.004665	0.000047
400	0.000309	0.000619	0.000006

Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML



Taylor's approximation & gradient descent in multiple variables

Multivariate calculus refresher



$$y = f(x_1, x_2, ...x_n) : \mathcal{R}^n \to \mathcal{R}$$

$$\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots \frac{\partial f}{\partial x_n}$$

- What is partial derivative?
- What is directional derivative (aka slope)?

Gradient: Direction of steepest ascent

Max value when theta = 0

Negative gradient: Direction of steepest descent

$$D_{\tilde{\mathbf{u}}} f(a,b) = (\nabla f)^T \vec{u}$$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1}\Big|_{(a,b)} \\ \frac{\partial f}{\partial x_2}\Big|_{(a,b)} \end{bmatrix}$$

Multivariate function approximation at any constant z

- f(x) is multivariate continuous, differentiable at x=z

$$p(x) \approx f(z) + f'(z)(x-z)$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ .. \\ x_n \end{bmatrix} \quad z = \begin{bmatrix} z_1 \\ z_2 \\ .. \\ z_n \end{bmatrix}$$

$$p(x) \approx f(z) + \frac{\partial f(z)}{\partial x_1}(x-z) + \frac{\partial f(z)}{\partial x_2}(x-z) + .. + \frac{\partial f(z)}{\partial x_n}(x-z)$$

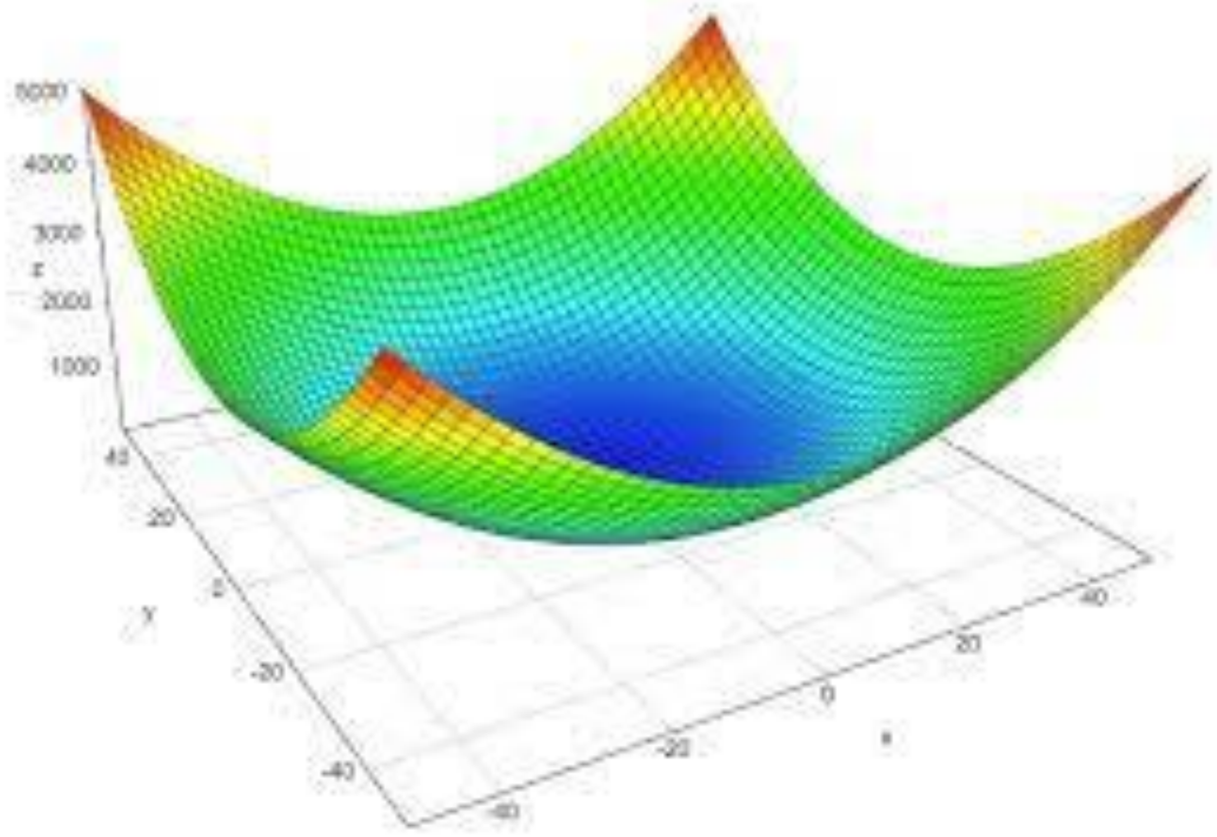
$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ .. \end{bmatrix}$$

$$p(x) \approx f(z) + \nabla f(z)^T (x-z)$$

You have to remember this

Multivariate gradient descent with first order approx.

$$p(x) \approx f(z) + \nabla f(z)^T (x - z)$$



$$\mathcal{L}(\mathbf{w}) \approx \mathcal{L}(\mathbf{z}) + \nabla \mathcal{L}(\mathbf{z})^T (\mathbf{w} - \mathbf{z})$$

$$w \rightarrow z; \qquad w - z = s$$

$$\mathcal{L}(\underbrace{z + s}_{w_{new}}) \approx \mathcal{L}(\underbrace{z}_{w_{old}}) + \nabla \mathcal{L}(\underbrace{z}_{w_{old}})^T s$$

$$s = -\alpha \nabla \mathcal{L}(w)$$

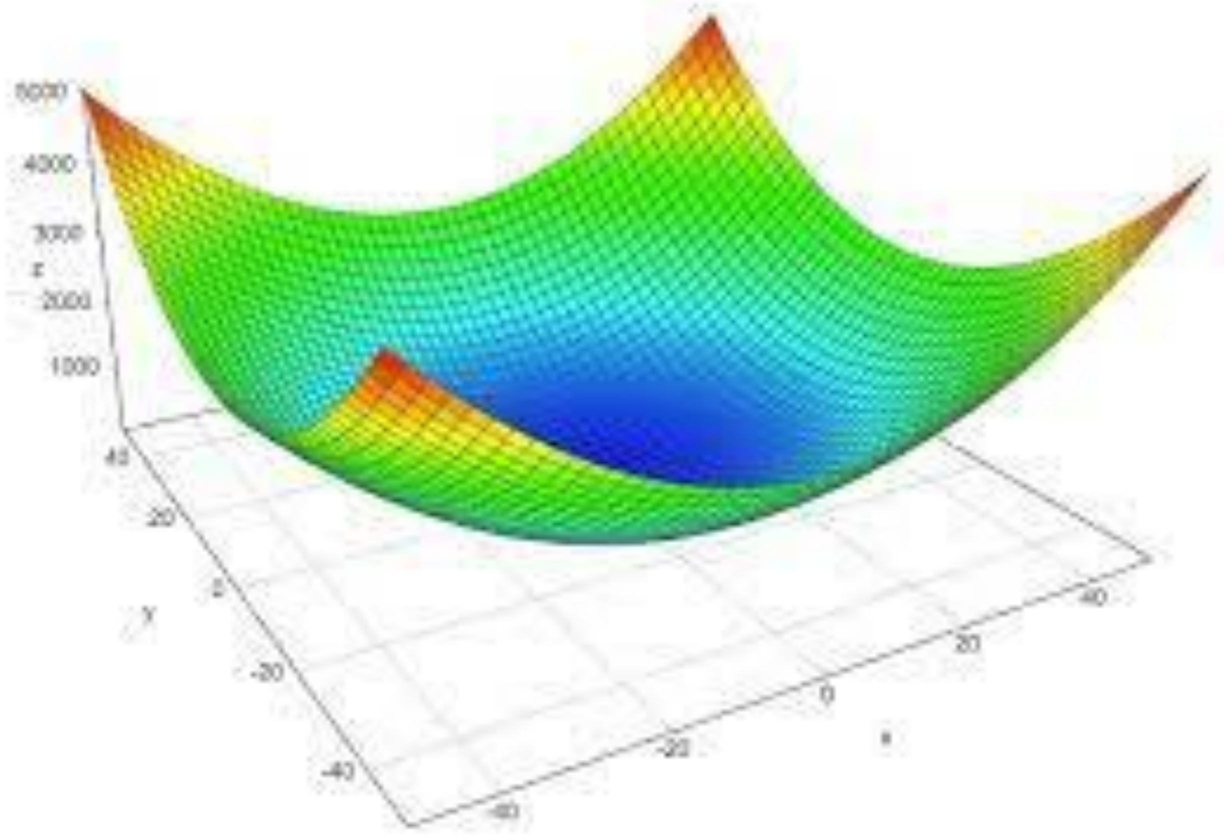
$$w_{new} = w_{old} - \alpha \nabla \mathcal{L}(w_{old})$$

$$\mathcal{L}(w_{new}) \approx \mathcal{L}(w_{old}) - \alpha \nabla \mathcal{L}(w_{old})^T \nabla \mathcal{L}(w_{old})$$

Newton's method: second order approximation

$$p(x) \approx f(z) + f'(z)x + \frac{1}{2}f''(z)(x-z)^2$$

$$p(x) \approx f(z) + \nabla f(z)^T(x-z) + \frac{1}{2}(w-z)^T\mathbf{H}(w-z)$$



$$\mathcal{L}(\mathbf{w}) \approx \mathcal{L}(\mathbf{z}) + \nabla\mathcal{L}(\mathbf{z})^T(\mathbf{w}-\mathbf{z}) + \frac{1}{2}(w-z)^T\mathbf{H}(w-z)$$

$$w \to z; \qquad w - z = s$$

$$H(\mathbf{w}) = \begin{pmatrix} \frac{\partial^2 \ell}{\partial w_1^2} & \frac{\partial^2 \ell}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 \ell}{\partial w_1 \partial w_n} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{\partial^2 \ell}{\partial w_n \partial w_1} & \cdots & \cdots & \frac{\partial^2 \ell}{\partial w_n^2} \end{pmatrix},$$


QUESTIONS

Literature Survey