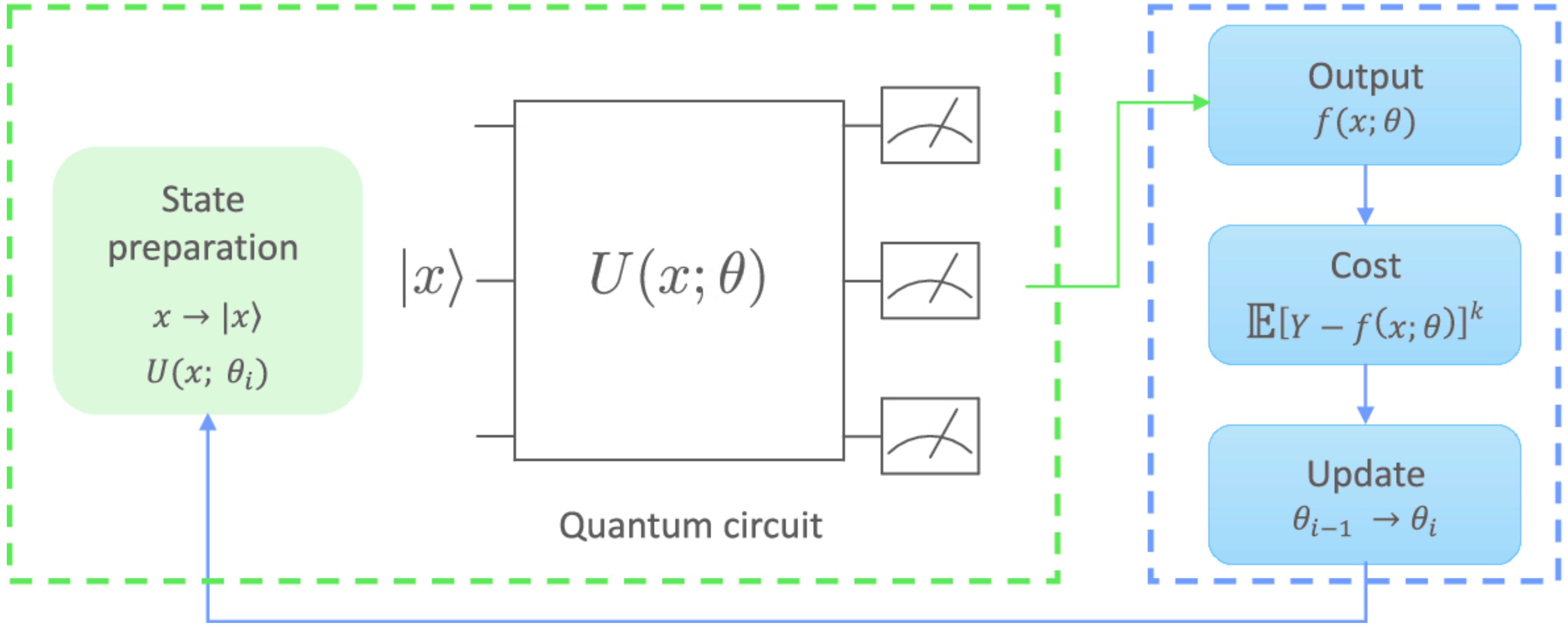
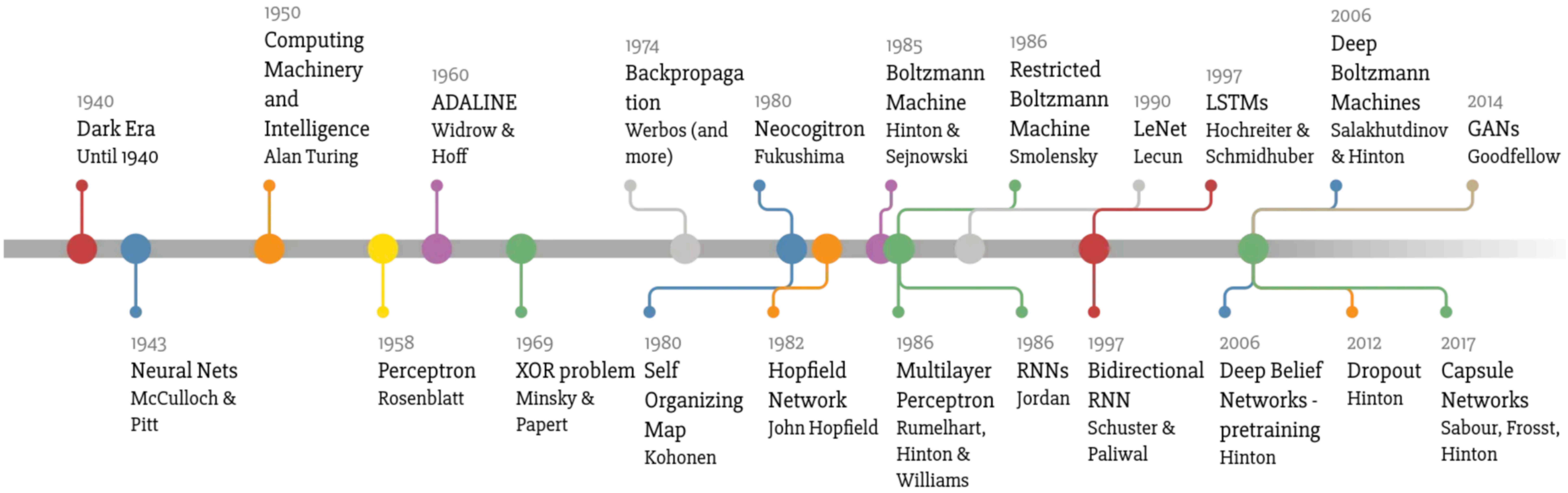
Training quantum models at scale

Amira Abbas,
University of KwaZulu-Natal



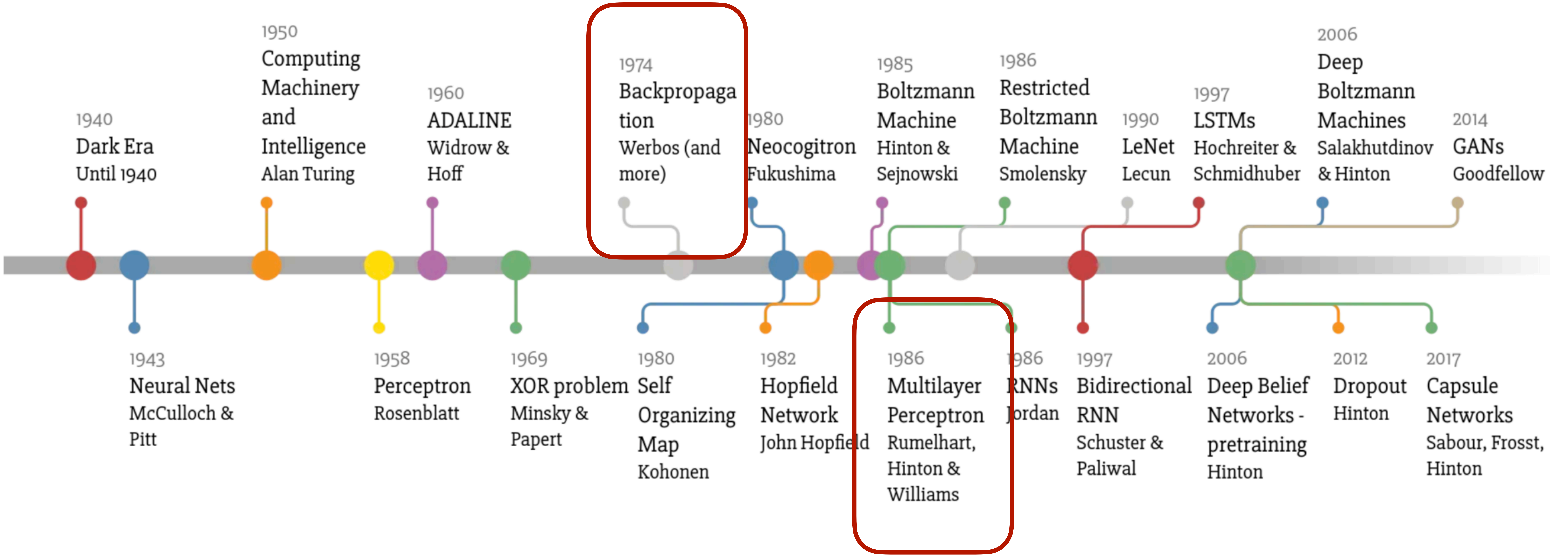
UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Deep Learning Timeline



Made by Favio Vázquez

Deep Learning Timeline



Made by Favio Vázquez

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

~175 billion parameters

A

Tell me in 2 sentences why backpropagation is so important.

Backpropagation is important because it allows artificial neural networks to learn and improve their performance over time through the adjustment of their weights. Without backpropagation, training complex neural networks with many layers and connections would be impractical, if not impossible.

chatGPT

How does backpropagation work?

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

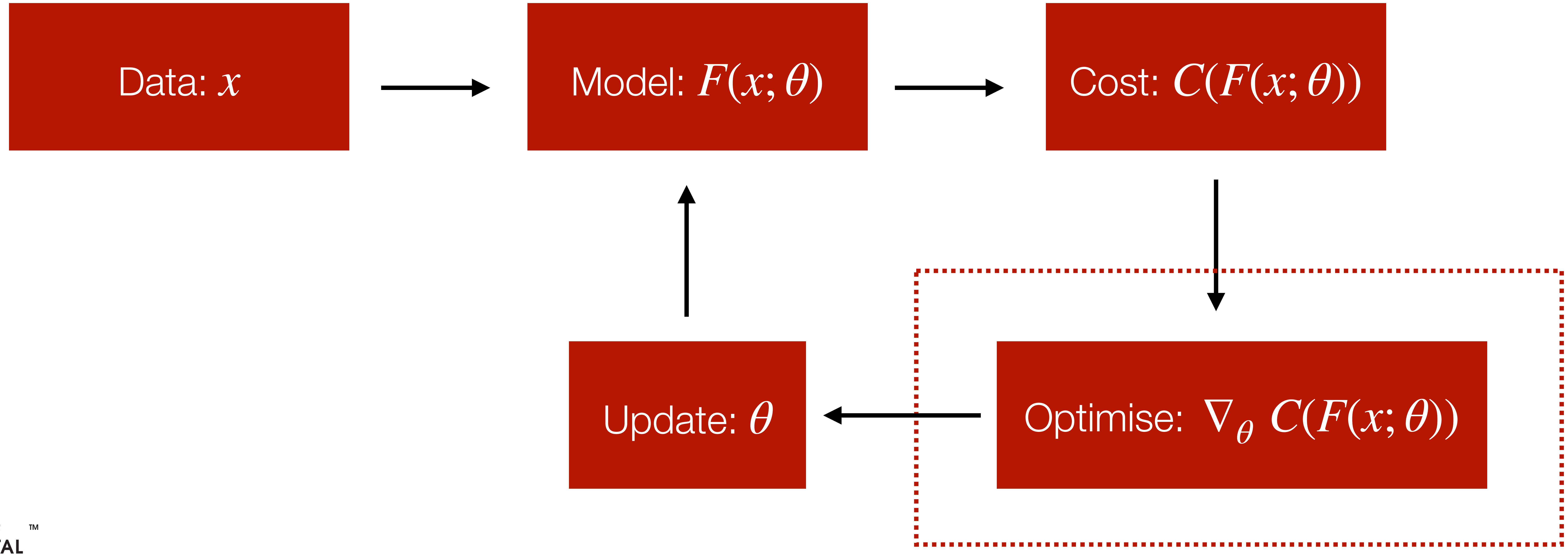
How does backpropagation work?

- A recipe to compute gradients

UNIVERSITY OF KWAZULU-NATAL INYUVESI YAKWAZULU-NATALI

How does backpropagation work?

- A recipe to compute gradients



How does backpropagation work?

- A recipe to compute gradients
- The first computationally efficient method update parameters of a neural network

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

How does backpropagation work?

- A recipe to compute gradients
- The first computationally efficient method update parameters of a neural network

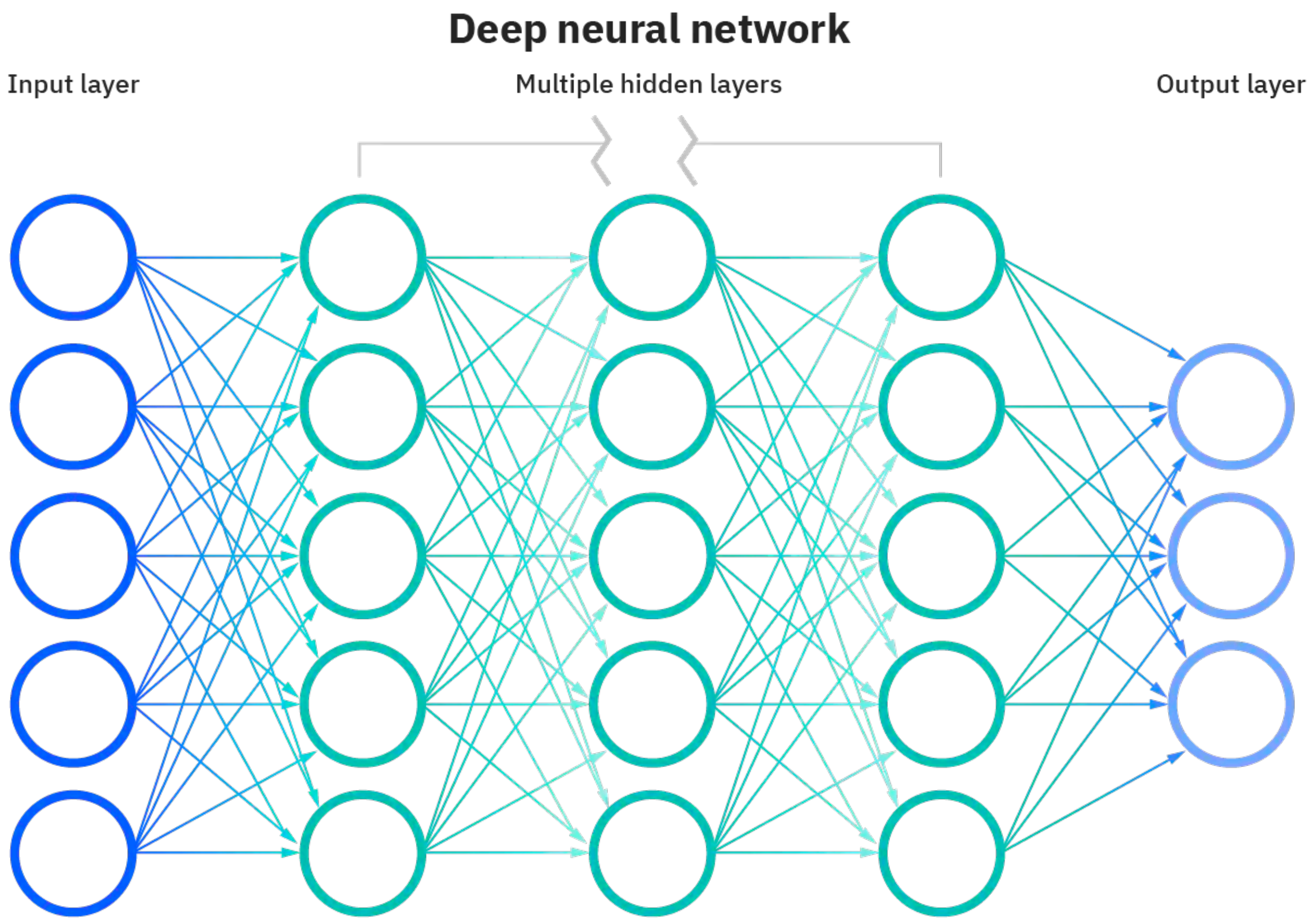
$$f(X;\vec{\theta}) = \sigma(\theta_L(\sigma(\theta_{L-1}...\theta_1(X))))$$

- Often solely attributed to the chain rule

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

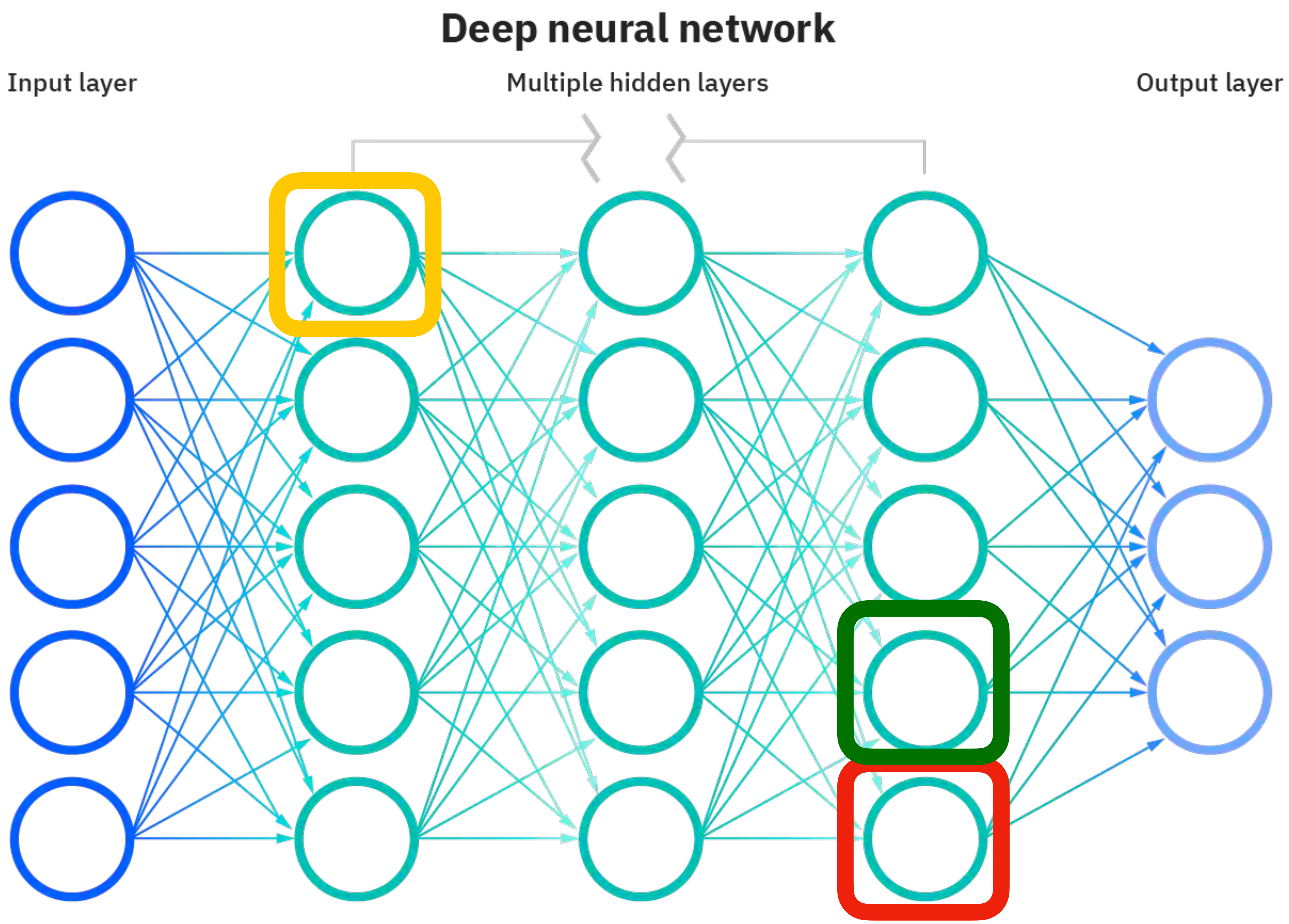
How does backpropagation work?

$$f(X;\vec{\theta}) = \sigma(\theta_L(\sigma(\theta_{L-1}...\theta_1(X))))$$



How does backpropagation work?

$$f(X;\vec{\theta}) = \sigma(\theta_L(\sigma(\theta_{L-1}...\theta_1(X))))$$



$$\frac{\partial f(X;\vec{\theta})}{\partial \theta_1}$$

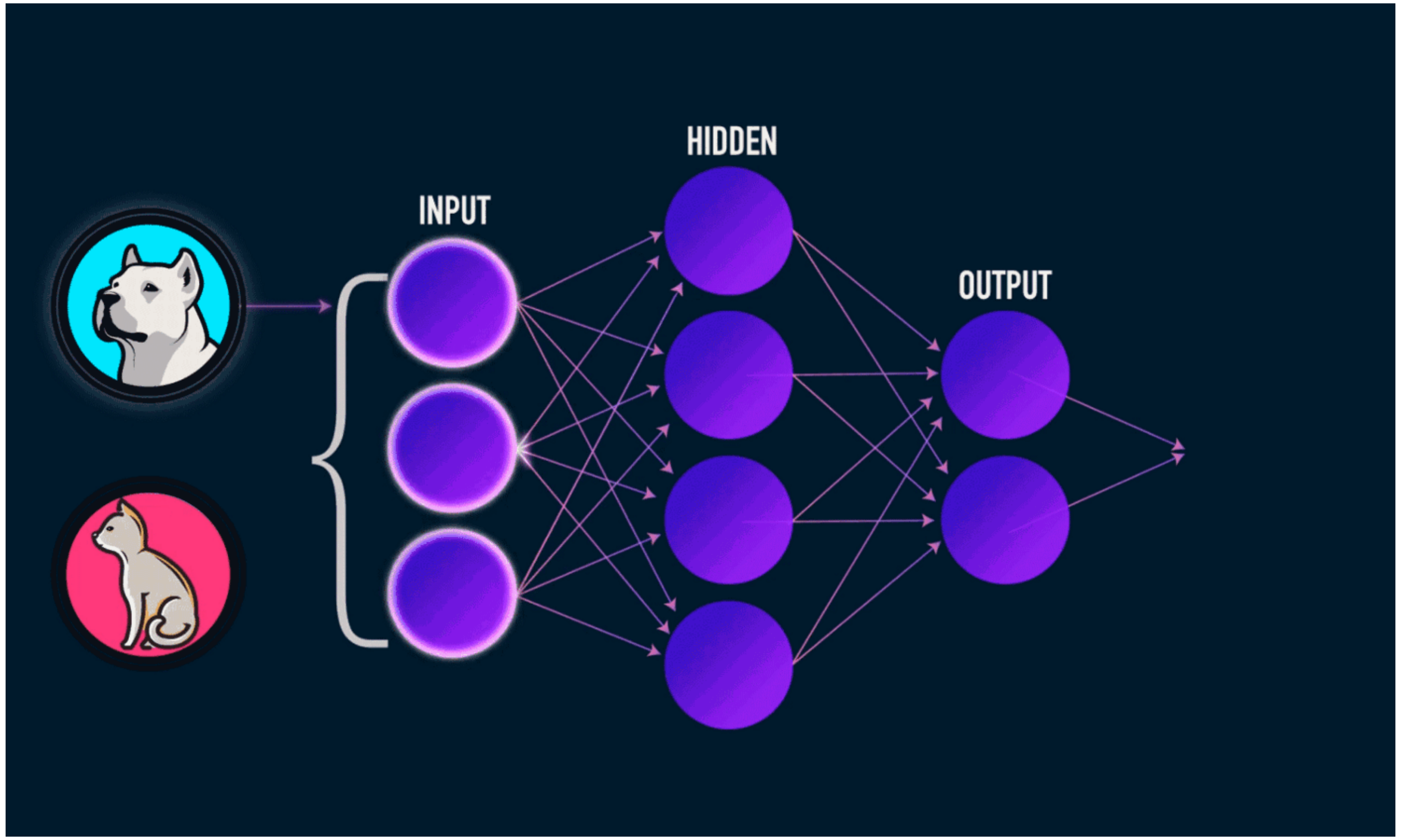
$$\vdots$$

$$\frac{\partial f(X;\vec{\theta})}{\partial \theta_L - 1}$$

$$\frac{\partial f(X;\vec{\theta})}{\partial \theta_L}$$

How does backpropagation work?

- As a neural network function is being computed, intermediate information is cleverly stored and reused for gradient computation - dynamic programming



UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Memory and time

Memory and time

- Neural network with M parameters $F(\theta),\ \theta \in \mathbb{R}^M$

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Memory and time

- Neural network with M parameters $F(\theta),\ \theta \in \mathbb{R}^M$
- Cost to compute the function in time: $\text{TIME}(F(\theta))$
- Cost to compute the function in memory: $\text{MEMORY}(F(\theta))$

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

Memory and time

- Neural network with M parameters $F(\theta), \ \theta \in \mathbb{R}^M$
- Cost to compute the function in time: $\mathrm{TIME}(F(\theta))$
- Cost to compute the function in memory: $\mathrm{MEMORY}(F(\theta))$

$$\mathrm{TIME}(\nabla F(\theta)) \leq \omega_1 \ \mathrm{TIME}(F(\theta))$$
$$\mathrm{MEMORY}(\nabla F(\theta)) \leq \omega_2 \ \mathrm{MEMORY}(F(\theta))$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Memory and time

- Neural network with M parameters $F(\theta), \ \theta \in \mathbb{R}^M$
- Cost to compute the function in time: $\mathrm{TIME}(F(\theta))$
- Cost to compute the function in memory: $\mathrm{MEMORY}(F(\theta))$

$$\mathrm{TIME}(\nabla F(\theta)) \leq \omega_1 \ \mathrm{TIME}(F(\theta))$$
$$\mathrm{MEMORY}(\nabla F(\theta)) \leq \omega_2 \ \mathrm{MEMORY}(F(\theta))$$

$$\omega_1, \omega_2 \in [2, 4]$$

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

An example: naive gradient computation

- Neural network: M = 1000 parameters

$$\text{TIME}(F(\theta)) = 0.01 \text{ seconds}$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

An example: naive gradient computation

- Neural network: M = 1000 parameters

$$\text{TIME}(F(\theta)) = 0.01 \text{ seconds}$$

$$\text{TIME}(\nabla F(\theta)) = 10 \text{ seconds}$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

An example: naive gradient computation

- Neural network: M = 1 billion parameters

$$\mathrm{TIME}(F(\theta)) = 60 \text{ seconds}$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

An example: naive gradient computation

- Neural network: M = 1 billion parameters

$$\text{TIME}(F(\theta)) = 60 \text{ seconds}$$

$$\text{TIME}(\nabla F(\theta)) \sim 31 \text{ years}$$

UNIVERSITY OF KWAZULU-NATAL™ INYUVESI YAKWAZULU-NATALI

An example: backpropagation scaling

- Neural network: M = 1 billion parameters

$$\text{TIME}(F(\theta)) = 60 \text{ seconds}$$

$$\text{TIME}(\nabla F(\theta)) \sim 5 \text{ minutes}$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Relative complexity

$$\text{TIME}(\nabla F(\theta)) \leq \omega_1 \ \text{TIME}(F(\theta))$$

$$\text{MEMORY}(\nabla F(\theta)) \leq \omega_2 \ \text{MEMORY}(F(\theta))$$

$$\omega_1, \omega_2 \in [2, 4]$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Quantum backpropagation?

$$\mathrm{TIME}(\nabla F(\theta)) \leq \omega_1 \ \mathrm{TIME}(F(\theta))$$

$$\mathrm{MEMORY}(\nabla F(\theta)) \leq \omega_2 \ \mathrm{MEMORY}(F(\theta))$$

$$\omega_1, \omega_2 \in [2, 4]$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Quantum backpropagation?

$$\mathrm{TIME}(\nabla F(\theta)) \leq \omega_1 \ \mathrm{TIME}(F(\theta))$$

$$\mathrm{MEMORY}(\nabla F(\theta)) \leq \omega_2 \ \mathrm{MEMORY}(F(\theta))$$

$$\omega_1, \omega_2 \in [2, 4]$$

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

Quantum backpropagation?

$$\mathrm{TIME}(\nabla F(\theta)) \leq \omega_1 \ \mathrm{TIME}(F(\theta)) \qquad \omega_1 = O(\log(M))$$

$$F(\theta) = \langle \psi(\theta)|O|\psi(\theta)\rangle$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Simple variational model

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Simple variational model

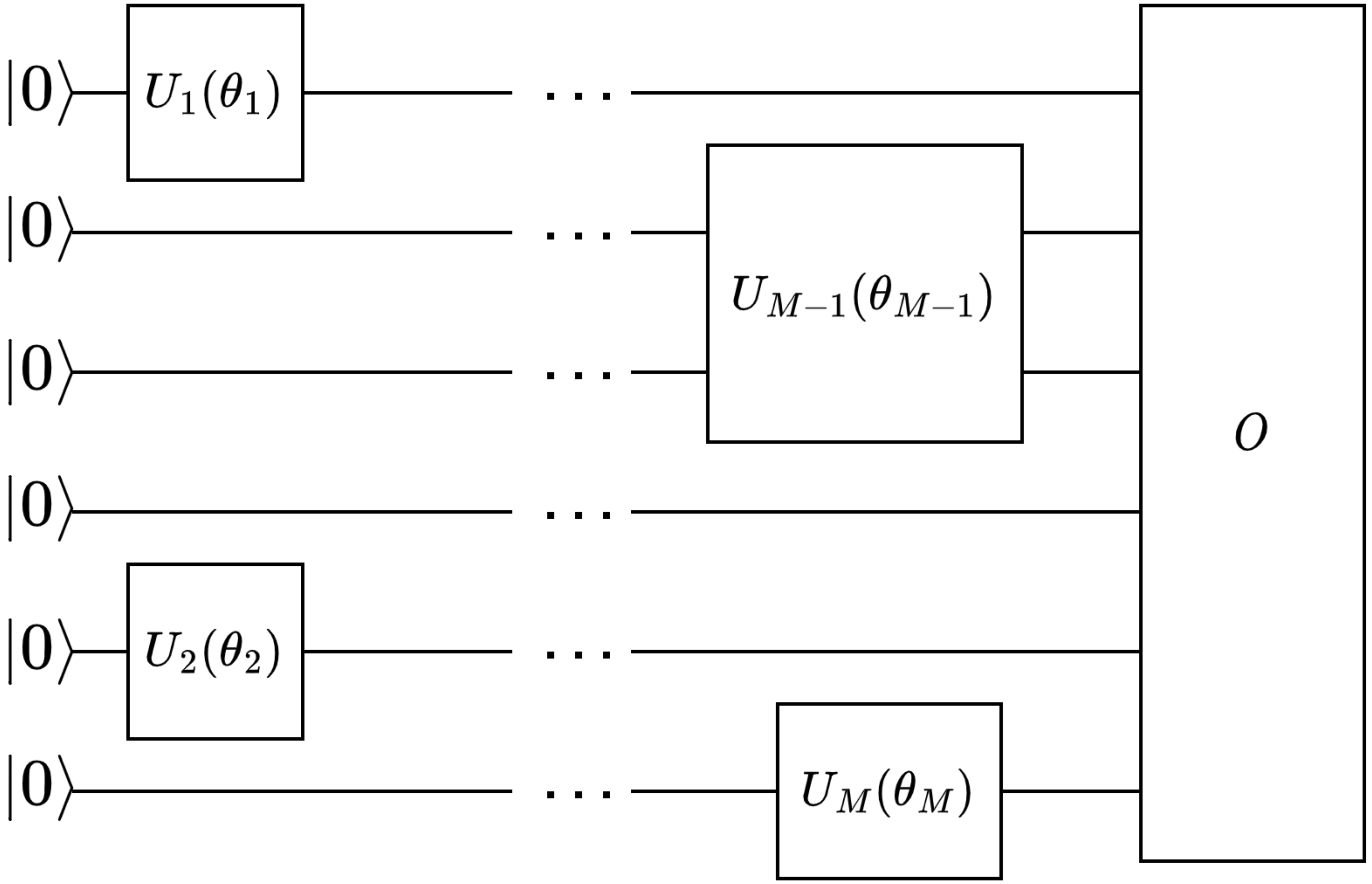
$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j) |0\rangle$$

Simple variational model

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle$$



UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Simple variational model

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

Simple variational model

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

$$\frac{\partial|\psi(\theta)\rangle}{\partial\theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j}(-iP_k)\prod_{l=1}^{k} e^{-i\theta_l P_l}|0\rangle$$

Simple variational model

$$F(\theta) = \langle \psi(\theta)|O|\psi(\theta)\rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

$$\frac{\partial|\psi(\theta)\rangle}{\partial\theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j}(-iP_k)\prod_{l=1}^{k} e^{-i\theta_l P_l}|0\rangle$$

$$[F'(\theta)]_{\theta_k} = -2\ \mathrm{Im}\ \langle\psi(\theta)|O\frac{\partial}{\partial\theta_k}|\psi(\theta)\rangle$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle\psi(\theta)|O|\psi(\theta)\rangle$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

$$\mathrm{TIME}(F(\theta)) = M$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

$$\text{TIME}(F(\theta)) = M/\epsilon^2$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$[F'(\theta)]_{\theta_k} = -2 \text{ Im } \langle \psi(\theta) | O \frac{\partial}{\partial \theta_k} | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j) |0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j} |0\rangle$$

$$\text{TIME}(F(\theta)) = M/\epsilon^2$$

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$[F'(\theta)]_{\theta_k} = -2 \text{ Im } \langle \psi(\theta) | O \frac{\partial}{\partial \theta_k} | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

$$\frac{\partial |\psi(\theta)\rangle}{\partial \theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j}(-iP_k)\prod_{l=1}^{k} e^{-i\theta_l P_l}|0\rangle$$

$$\text{TIME}(F(\theta)) = M/\epsilon^2$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$[F'(\theta)]_{\theta_k} = -2 \text{ Im } \langle \psi(\theta) | O \frac{\partial}{\partial \theta_k} | \psi(\theta) \rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j) |0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j} |0\rangle$$

$$\frac{\partial |\psi(\theta)\rangle}{\partial \theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j} (-iP_k) \prod_{l=1}^{k} e^{-i\theta_l P_l} |0\rangle$$

$$\text{TIME}(F(\theta)) = M/\epsilon^2$$

$$\text{TIME}([F'(\theta)]_{\theta_k}) = M/\epsilon^2$$

Naive quantum gradient scaling

- Unit cost for each parameterised unitary, of which there are M of them

$$F(\theta) = \langle\psi(\theta)|O|\psi(\theta)\rangle$$

$$[F'(\theta)]_{\theta_k} = -2\ \mathrm{Im}\ \langle\psi(\theta)|O\frac{\partial}{\partial\theta_k}|\psi(\theta)\rangle$$

$$|\psi(\theta)\rangle = \prod_{j=1}^{M} U_j(\theta_j)|0\rangle = \prod_{j=1}^{M} e^{-i\theta_j P_j}|0\rangle$$

$$\frac{\partial|\psi(\theta)\rangle}{\partial\theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j}(-iP_k)\prod_{l=1}^{k} e^{-i\theta_l P_l}|0\rangle$$

$$\mathrm{TIME}(F(\theta)) = M/\epsilon^2$$

$$\mathrm{TIME}([F'(\theta)]_{\theta_k}) = M/\epsilon^2$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

$$\mathrm{TIME}(\nabla F(\theta)) = M \cdot M/\epsilon^2 \;\; = M \cdot \mathrm{TIME}(\mathrm{F}(\theta))$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Naive quantum gradient scaling

$$\mathrm{TIME}(\nabla F(\theta)) = M \cdot M/\epsilon^2 \;\; = M \cdot \mathrm{TIME}(\mathrm{F}(\theta))$$

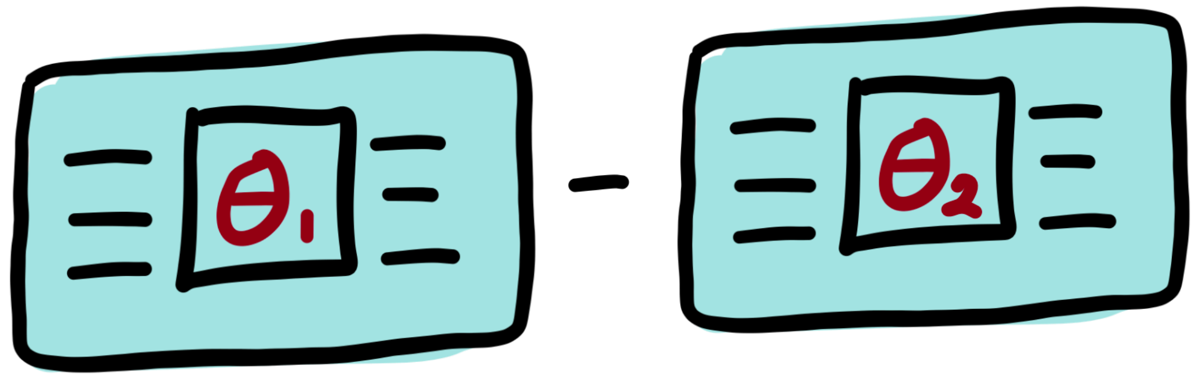
$$\mathrm{TIME}(\nabla \mathrm{F}(\theta)) = \log(M) \cdot \mathrm{TIME}(\mathrm{F}(\theta))$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Quantum backpropagation?

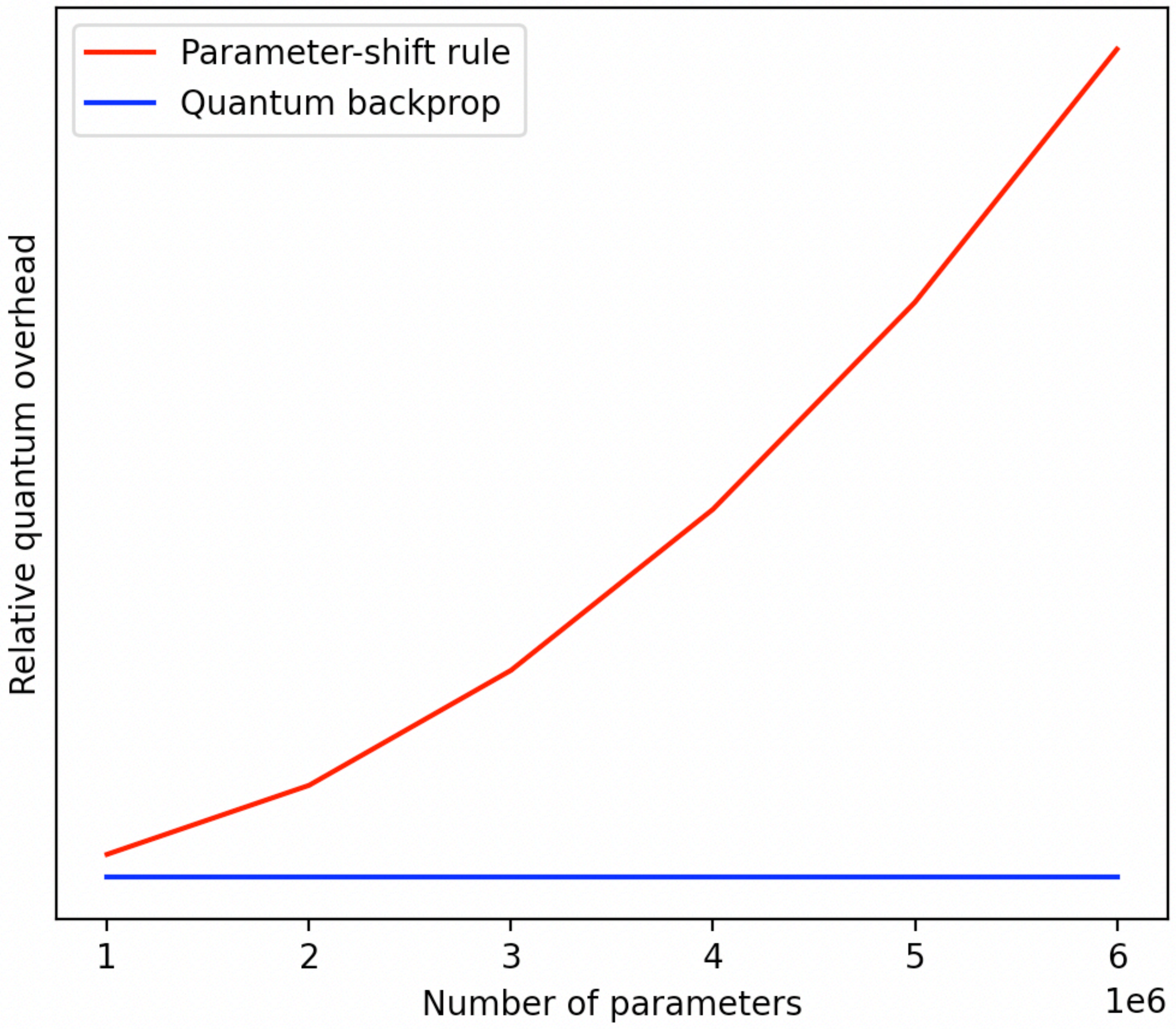
- Parameter shift rule does not yield backprop scaling

$$\nabla_\theta f = f(\theta_1) - f(\theta_2)$$



Quantum backpropagation?

- Parameter shift rule does not yield backprop scaling



Quantum backpropagation?

- Parameter shift rule does not yield backprop scaling
- SPSA (meant to be dimension independent) will fail

Quantum backpropagation?

- Parameter shift rule does not yield backprop scaling
- SPSA (meant to be dimension independent) will fail
- All *known* methods fail (unless special case models are considered)

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Is there something more fundamental preventing us from achieving backpropagation scaling?

$$[F'(\theta)]_{\theta_k} = -2 \text{ Im } \langle \psi(\theta) | O \frac{\partial}{\partial \theta_k} | \psi(\theta) \rangle$$

No cloning theorem

$$[F'(\theta)]_{\theta_k} = -2\ \mathrm{Im}\ \langle\psi(\theta)|O\frac{\partial}{\partial\theta_k}|\psi(\theta)\rangle$$

Measurement collapse

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Connecting the gradient problem to a more general open problem

Simplifying gradients

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

Simplifying gradients

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$\frac{\partial |\psi(\theta)\rangle}{\partial \theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j} (-iP_k) \prod_{l=1}^{k} e^{-i\theta_l P_l} |0\rangle$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Simplifying gradients

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$\frac{\partial |\psi(\theta)\rangle}{\partial \theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j} (-iP_k) \prod_{l=1}^{k} e^{-i\theta_l P_l} |0\rangle$$

$$[F'(\theta)]_{\theta_k} =$$

Simplifying gradients

$$F(\theta) = \langle \psi(\theta)|O|\psi(\theta)\rangle$$

$$\frac{\partial|\psi(\theta)\rangle}{\partial\theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j}(-iP_k)\prod_{l=1}^{k} e^{-i\theta_l P_l}|0\rangle$$

$$[F'(\theta)]_{\theta_k} = -2\ \mathrm{Im}\left[\langle 0|\left(\prod_{j=1}^{M} e^{i\theta_j P_j}\right) O \left(\prod_{m=k+1}^{M} e^{-i\theta_m P_m}\right)(-iP_k)\left(\prod_{l=1}^{k} e^{-i\theta_l P_l}\right)|0\rangle\right]$$

Simplifying gradients

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

$$\frac{\partial |\psi(\theta)\rangle}{\partial \theta_k} = \prod_{j=k+1}^{M} e^{-i\theta_j P_j} (-iP_k) \prod_{l=1}^{k} e^{-i\theta_l P_l} |0\rangle$$

$$[F'(\theta)]_{\theta_k} = -2 \ \mathrm{Im} \left[\langle 0| \left(\prod_{j=1}^{M} e^{i\theta_j P_j} \right) O \left(\prod_{m=k+1}^{M} e^{-i\theta_m P_m} \right) (-iP_k) \left(\prod_{l=1}^{k} e^{-i\theta_l P_l} \right) |0\rangle \right]$$

Let $\vec{\theta} = 0$ and $O = I$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Simplifying gradients

$$[F'(\theta)]_{\theta_k} = -2 \text{ Im} \left[\langle 0| \left(\prod_{j=1}^{M} e^{i\theta_j P_j} \right) O \left(\prod_{m=k+1}^{M} e^{-i\theta_m P_m} \right) (-iP_k) \left(\prod_{l=1}^{k} e^{-i\theta_l P_l} \right) |0\rangle \right]$$

Let $\vec{\theta} = 0$ and $O = I$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Simplifying gradients

$$[F'(\theta)]_{\theta_k} = -2\ \mathrm{Im}\left[\langle 0| \left(\prod_{j=1}^{M} e^{i\theta_j P_j}\right) O \left(\prod_{m=k+1}^{M} e^{-i\theta_m P_m}\right) (-iP_k) \left(\prod_{l=1}^{k} e^{-i\theta_l P_l}\right) |0\rangle\right]$$

Let $\vec{\theta} = 0$ and $O = I$

$$[F'(\theta)]_{\theta_k} = 2\ \mathrm{Re}\ \langle 0|P_k|0\rangle$$

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Simplifying gradients

$$[F'(\theta)]_{\theta_k} = -2 \ \mathrm{Im} \left[\langle 0| \left(\prod_{j=1}^{M} e^{i\theta_j P_j} \right) O \left(\prod_{m=k+1}^{M} e^{-i\theta_m P_m} \right) (-iP_k) \left(\prod_{l=1}^{k} e^{-i\theta_l P_l} \right) |0\rangle \right]$$

Let $\vec{\theta} = 0$ and $O = I$

$$[F'(\theta)]_{\theta_k} = 2 \ \mathrm{Re} \ \langle 0|P_k|0\rangle$$

Task: Estimate the above expectation value **for all** $k = 1, \ldots, M$ using as little resources as possible

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Shadow tomography

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Shadow Tomography of Quantum States*

Scott Aaronson†

Problem 1 (Shadow Tomography) *Given an unknown D-dimensional quantum mixed state ρ, as well as known 2-outcome measurements $E_1, \ldots, E_M$, each of which accepts ρ with probability $\operatorname{Tr}(E_i \rho)$ and rejects ρ with probability $1 - \operatorname{Tr}(E_i \rho)$, output numbers $b_1, \ldots, b_M \in [0, 1]$ such that $|b_i - \operatorname{Tr}(E_i \rho)| \leq \varepsilon$ for all i, with success probability at least $1 - \delta$. Do this via a measurement of $\rho^{\otimes k}$, where $k = k(D, M, \varepsilon, \delta)$ is as small as possible.*

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

Reusing quantum states

Reusing quantum states

- Partially destructive measurements

$$||\rho - \rho'||_{\mathrm{tr}}$$

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Reusing quantum states

- Partially destructive measurements

$$||\rho - \rho'||_{\mathrm{tr}} \leq \alpha$$

Reusing quantum states

- Partially destructive measurements

$$||\rho - \rho'||_{\mathrm{tr}} \leq \alpha$$

- Gentle measurements

Can we use shadow tomography for gradients?

Yes… *and no.*

Shadow tomography

$$\mathrm{TIME}(\nabla \mathrm{F}(\theta)) = \log(M) \cdot \mathrm{TIME}(\mathrm{F}(\theta))$$

Shadow tomography

$$\mathrm{TIME}(\nabla \mathrm{F}(\theta)) = \log(M) \cdot \mathrm{TIME}(\mathrm{F}(\theta))$$

- Leading techniques require storage of exponentially large offline models

Shadow tomography

$$\mathrm{TIME}(\nabla \mathrm{F}(\theta)) = \log(M) \cdot \mathrm{TIME}(\mathrm{F}(\theta))$$

- Leading techniques require storage of exponentially large offline models

$$\mathrm{MEMORY}(\nabla F(\theta)) \neq \omega \ \mathrm{MEMORY}(F(\theta))$$

Recap

Recap

- Information reuse in quantum models is not easy and is an inhibitor of true backprop scaling

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Recap

- Information reuse in quantum models is not easy and is an inhibitor of true backprop scaling

- Current gradient methods for quantum backprop do not achieve the desired scaling of resources

Closing remarks

Closing remarks

- There may be some restricted settings, perhaps where models are not as universal or powerful, where backprop scaling is attainable

Closing remarks

- There may be some restricted settings, perhaps where models are not as universal or powerful, where backprop scaling is attainable

- Shadow tomography: true computational complexity is still unknown

Closing remarks

- There may be some restricted settings, perhaps where models are not as universal or powerful, where backprop scaling is attainable
- Shadow tomography: true computational complexity is still unknown
- Is there a more general computational argument to rule out backprop?

Closing remarks

- There may be some restricted settings, perhaps where models are not as universal or powerful, where backprop scaling is attainable
- Shadow tomography: true computational complexity is still unknown
- Is there a more general computational argument to rule out backprop?
- New models or methods for optimisation? — If QML is to complete with classical ML