What are Barren Plateaus?

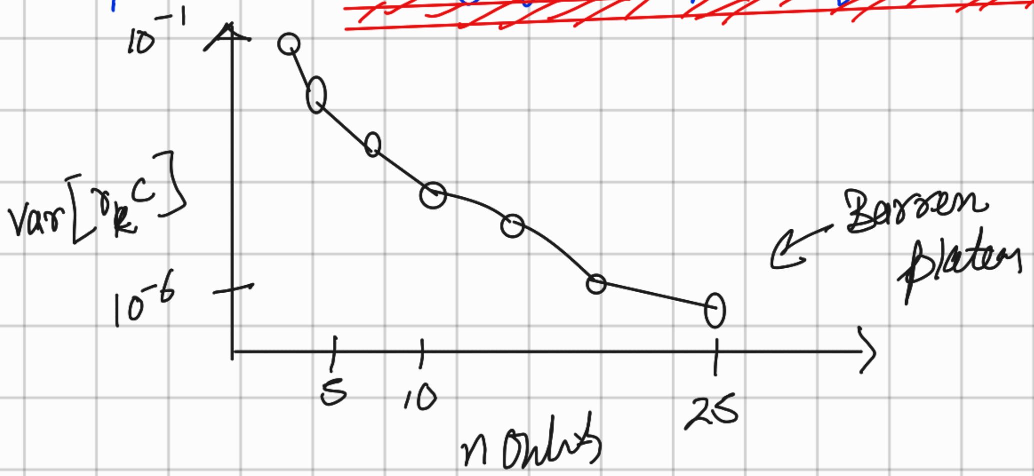
$$Var[\partial_k C] \sim \frac{1}{2^n} \quad + \quad P(|\partial_k C| \geq \delta) \leq \frac{Var[\partial_k C]}{\delta^2}$$

(Combining with Chebyshev ineq)

Probability of non-zero gradients vanishes exponentially with problem size.

and

shot required for training grows exponentially with problem size.

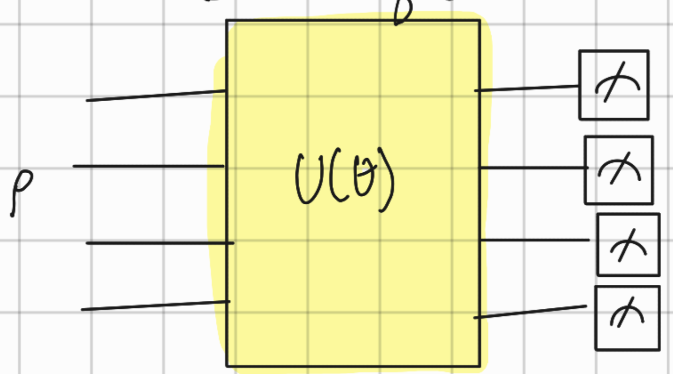


Trainability of a QML model!

While working on a QML model, we deal with a loss function that helps the optimizer train the quantum circuit better.

$$l_\theta(\rho, O) = Tr[U(\theta)\rho U^\dagger(\theta) O]$$

Sources of Barren Plateaus.



studying loss concentration (variance of loss) is equivalent of studying partial derivative concentration.

No Barren Plateaus	Barren Plateau
$Var_\theta[l_\theta(\rho, O)] \in \Omega(1/$	$Var_\theta[l_\theta(\rho, O)] \in O(\frac{1}{b^n})$ with $b > 1$

$\left(\frac{\ }{\text{poly}(n)}\right)$

Variance of the loss function decays atmost polynomially with number of qubits.

Variance will be exponentially smaller.

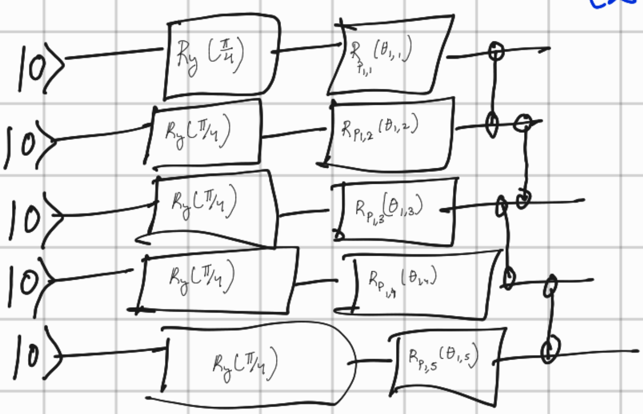
Here polynomial number of shots are required.

Here exponential number of shots are required

wellknown
Most common source of Barren Plateau

Expressiveness in the Circuit

(2-design)



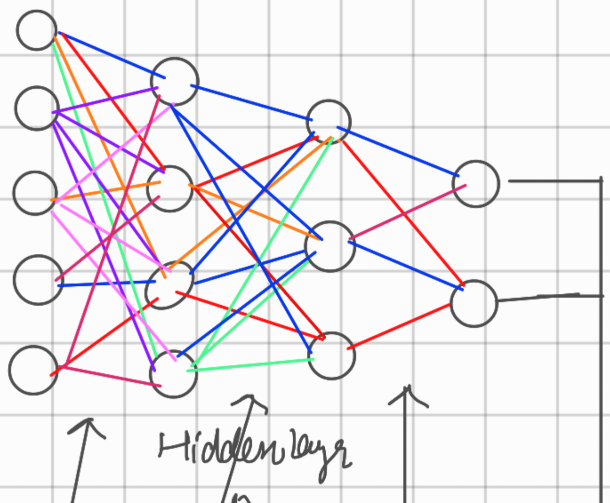
$$\mathrm{Var}_\theta\left[\ell_\theta(\rho, O)\right] = \frac{1}{4^n - 1}\left(\mathrm{Tr}[\rho^2] - \frac{1}{2^n}\right) \times \left(\mathrm{Tr}[O^2] - \frac{\mathrm{Tr}[O]^2}{2^n}\right)$$

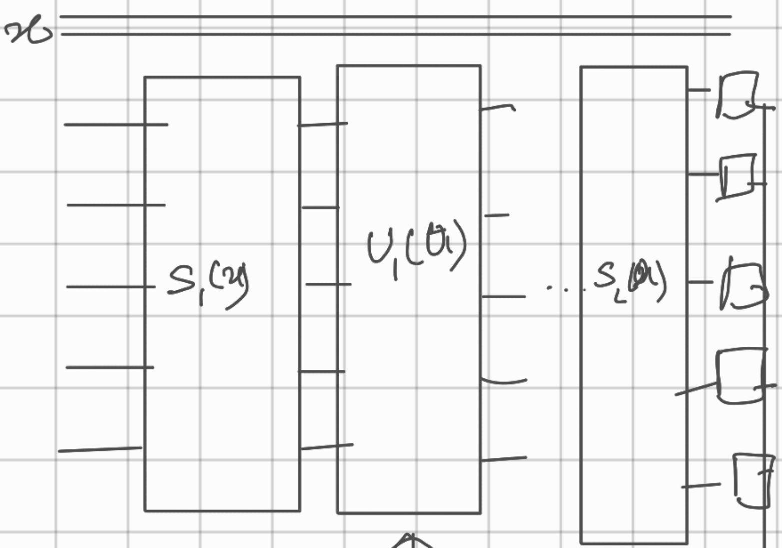
Barren Plateaus, Trainability Issues and How to avoid them:

Classical feed forward design

Quantum neural network.

Input layer





Information $L(x;\theta)$

Training

$$f_\theta(x) = \sigma(w_\theta x + b_\theta)$$

$$y_{nn}(x) = f_{\theta_L} \circ f_{\theta_{L-1}} \cdots \circ f_{\theta_1}(x)$$

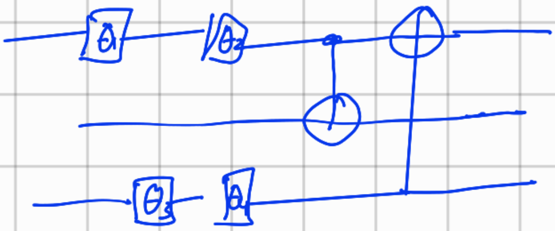
$$y_{qnn}(x) = \langle 0| U^\dagger M U |0\rangle$$

where

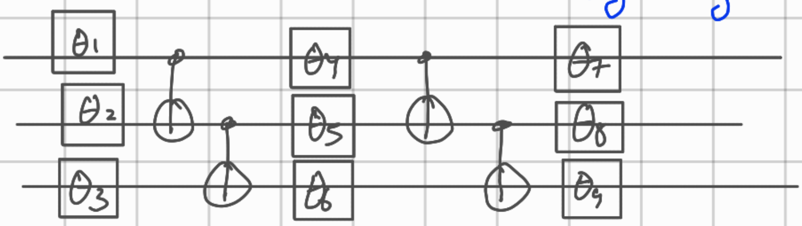
$$U(x;\theta) = \prod U_i(\theta_i) S_i(x)$$

Similar layered structure but different information flow.

Solving problems with parameterized quantum circuits:

Problem →  → Solution

Educated guess ("Expressivity and Entangibility")
Eg: UCC circuit for Quantum Chemistry
but it's not always possible. so we mostly link with Hardware efficient ansatz.
Eg: R_y-CNOT gate

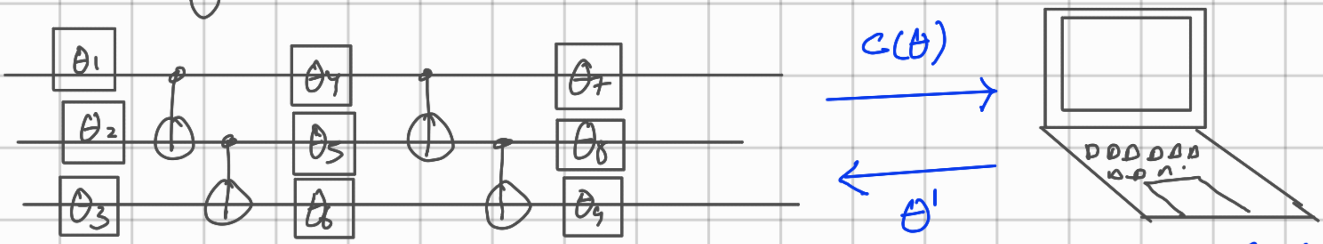


This ansatz has the advantage that you need to know very little about your problem and this ansatz will still work. But won't be much efficient or it may be the case that

the hidden structure of the symmetries that we get from the problem data or may want to use to build an ansatz are precisely part of the solution.

ref(Lie Algebra inspired Ansatz)

Vanishing Gradients



classical routine

$$C(\theta) = \langle 0 | U^{\dagger}(\theta)\, O\, U(\theta) | 0 \rangle$$

cost function for the optimization problem

gradients = derivatives of the cost functions with respect to parameters.

our ansatz is a deep randomized circuit where the depth of the circuit $d \sim O(\text{poly}(n))$ where $n \sim$ number of qubits. we have to use these many because we need to explore much of the Hilbert space, and there's Random initialization of the parameters, since we have no prior insight on where to start.

These two conditions are sufficient for the issue of vanishing gradients to arise.

what we observe is that the gradients of the cost function vanish

$$\langle \partial_k C \rangle = 0$$

and the variance of the gradients will be exponentially small with the number of qubits that you use.

$$\boxed{\mathrm{Var}[\partial_k C] \simeq 2^{-n}}$$

where n = number of qubits

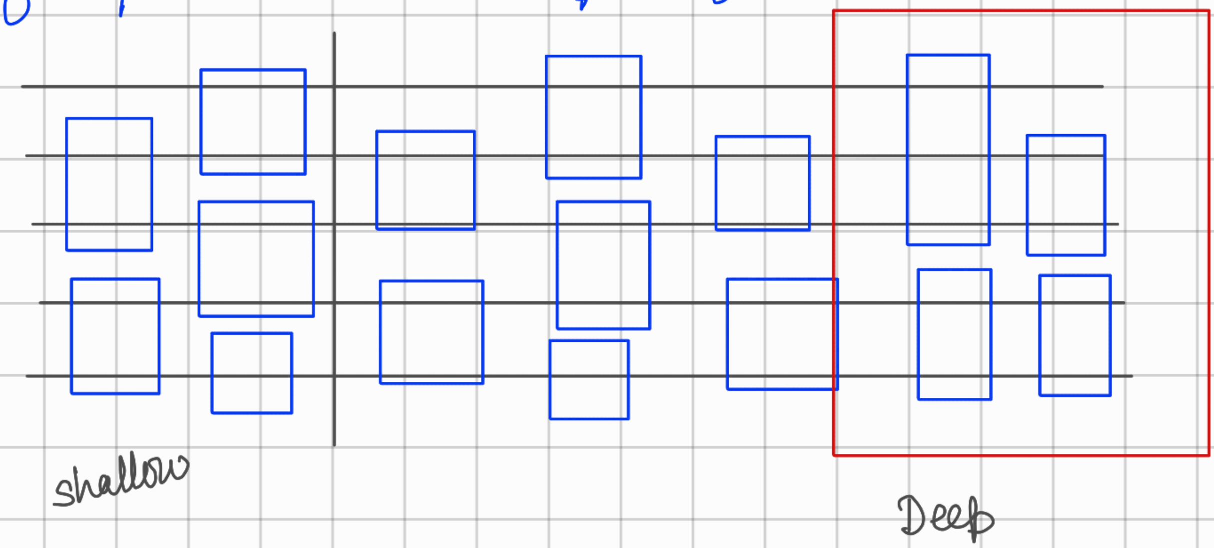
Gradients vanish exponentially with the number of qubits.

Hence the emergence of the Barren Plateau

Random quantum circuits quickly approximate the 2-design

Brief History of Barren Plateaus

The paper by J. McLean et al. Nature Commun (2018). showed that if your quantum circuit is deep enough:



and by deep it means that essentially they contain a number of layers which is a polynomial function of the number of qubits.

$$\text{Depth} \cong O(n^{1/L}) \quad \text{on } L\text{-dim arrays}$$

$$c(\theta) = \langle 0 | U^\dagger(\theta)\, O\, U(\theta) | 0 \rangle$$

where $O = c_0 \mathbb{I} + \sum_i \otimes_j c_i O_{ij}$