Concentration of Data Encoding in Parameterized Quantum Circuits

Guangxi Li[1,2], Ruilin Ye[2,3], Xuanqiang Zhao[2], Xin Wang[2]

[1] University of Technology Sydney, NSW, Australia

[2] Institute for Quantum Computing, Baidu Research, Beijing, China

[3] Peking University, Beijing, China

Oct 21, 2022

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Baidu Research

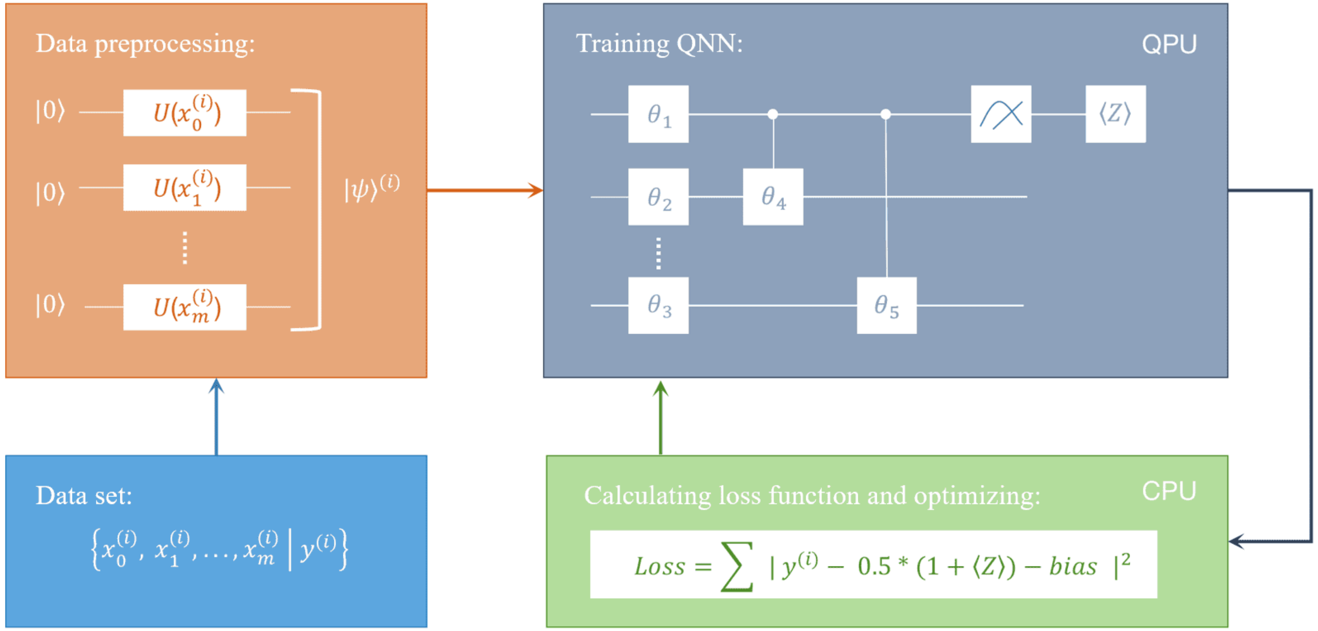
Hybrid Quantum-Classical Computing

Hybrid Quantum-Classical Computing

- A promising framework to achieve quantum advantages on current *noisy intermediate-scale quantum* (NISQ) devices

Hybrid Quantum-Classical Computing

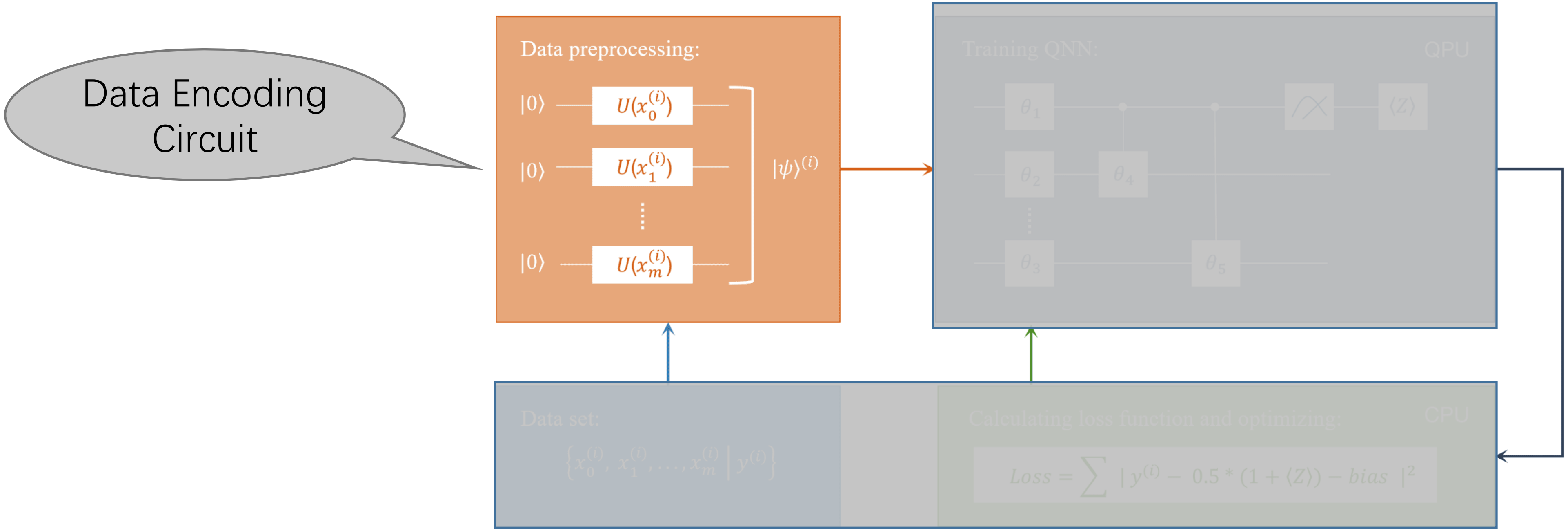
- A promising framework to achieve quantum advantages on current *noisy intermediate-scale quantum* (NISQ) devices



Flow chart of binary quantum classifier training from Paddle-Quantum

Hybrid Quantum-Classical Computing

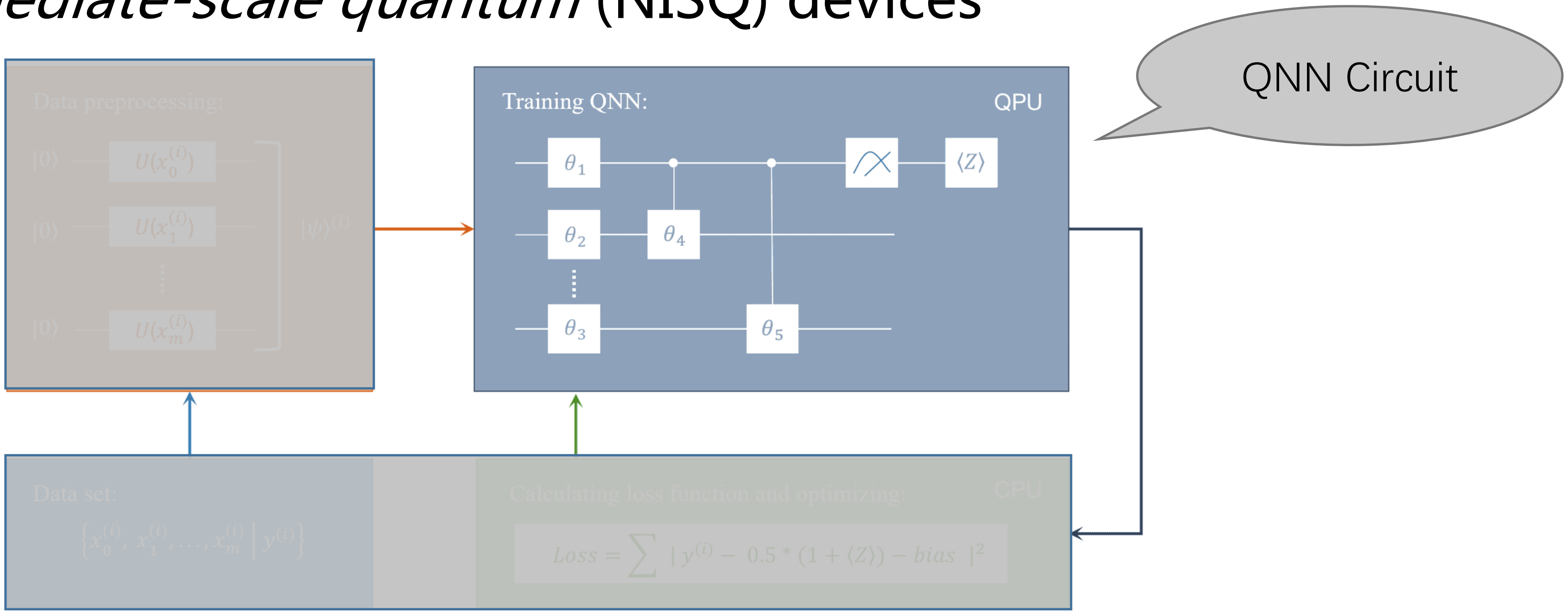
- A promising framework to achieve quantum advantages on current *noisy intermediate-scale quantum* (NISQ) devices



Flow chart of binary quantum classifier training from Paddle-Quantum

Hybrid Quantum-Classical Computing

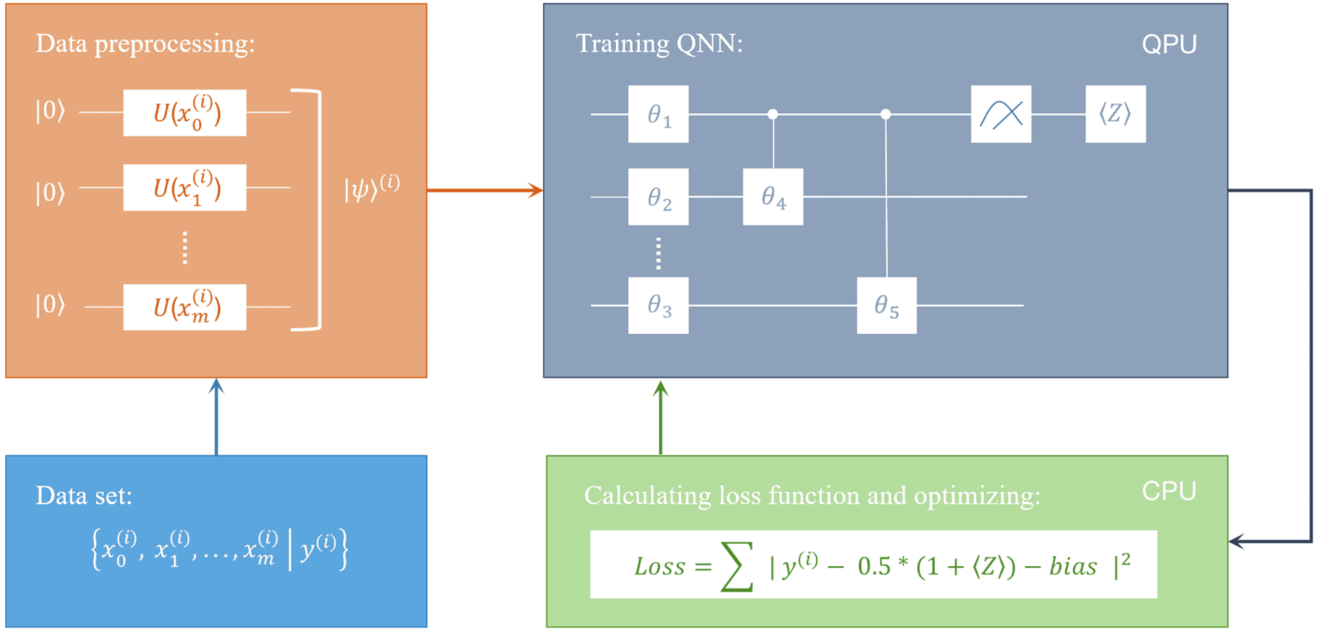
- A promising framework to achieve quantum advantages on current *noisy intermediate-scale quantum* (NISQ) devices



Flow chart of binary quantum classifier training from Paddle-Quantum

Hybrid Quantum-Classical Computing

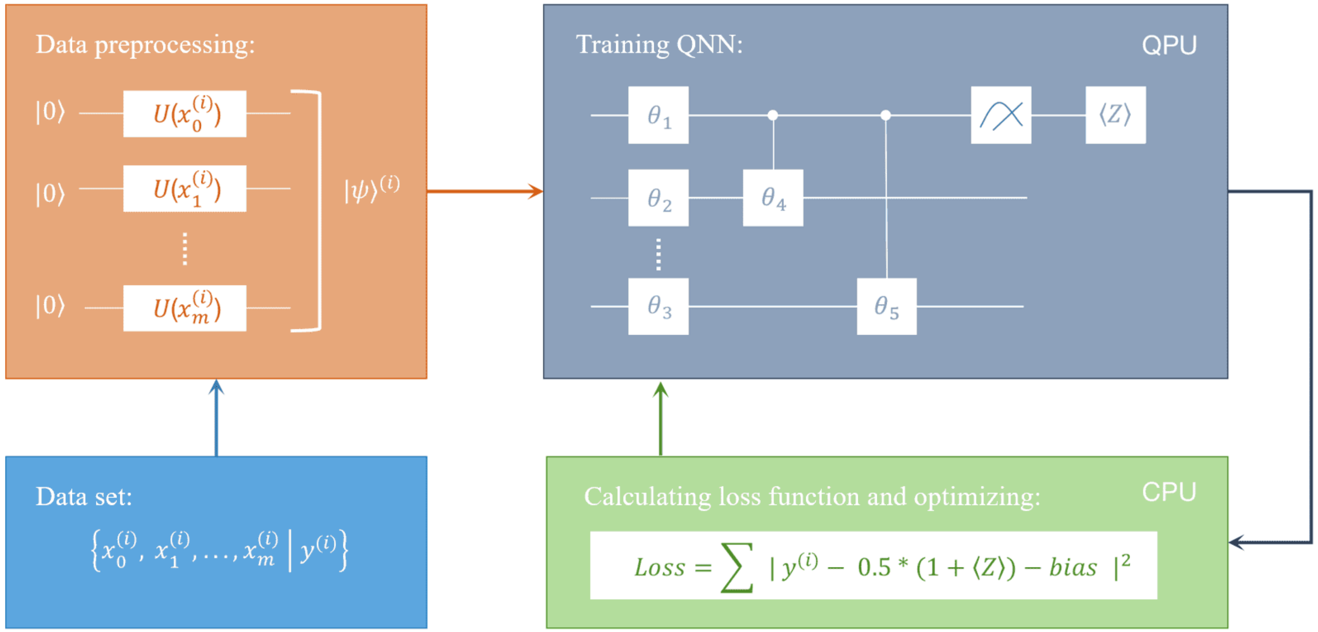
- A promising framework to achieve quantum advantages on current *noisy intermediate-scale quantum* (NISQ) devices



Flow chart of binary quantum classifier training from Paddle-Quantum

Hybrid Quantum-Classical Computing

- A promising framework to achieve quantum advantages on current *noisy intermediate-scale quantum* (NISQ) devices

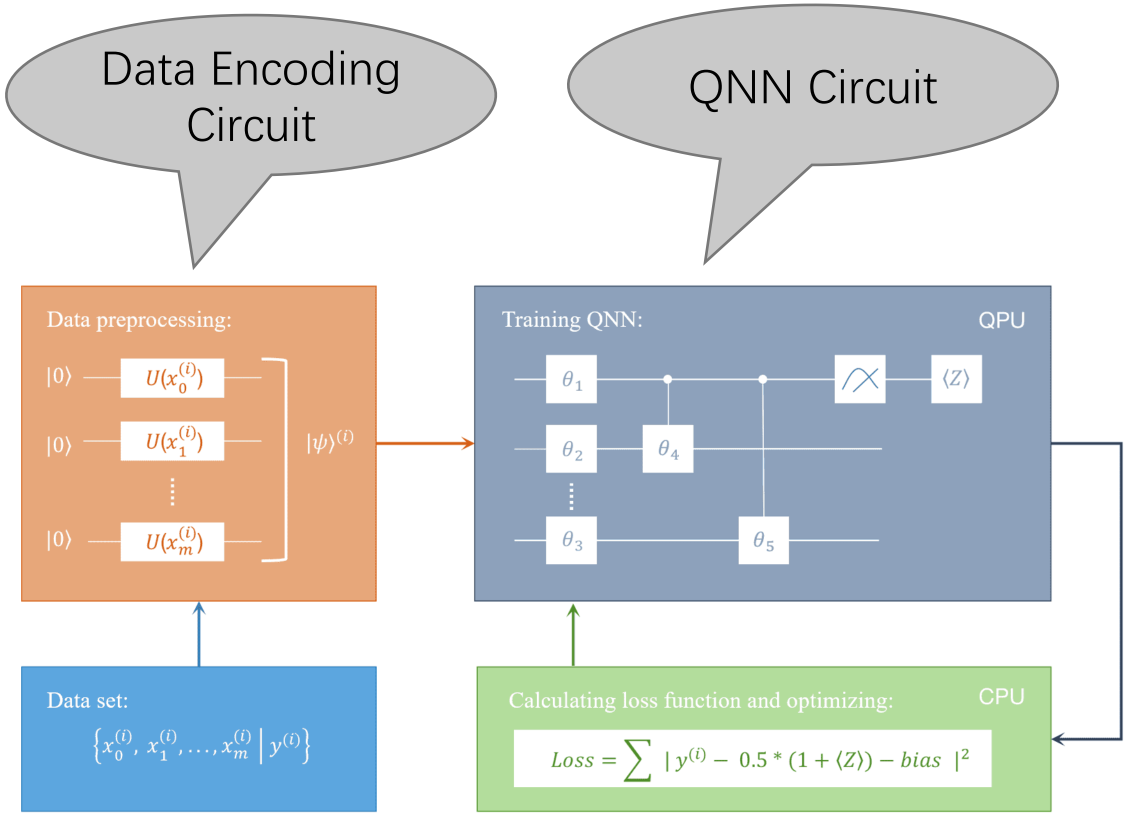


Flow chart of binary quantum classifier training from Paddle-Quantum

- The performance highly relies on the power of PQCs

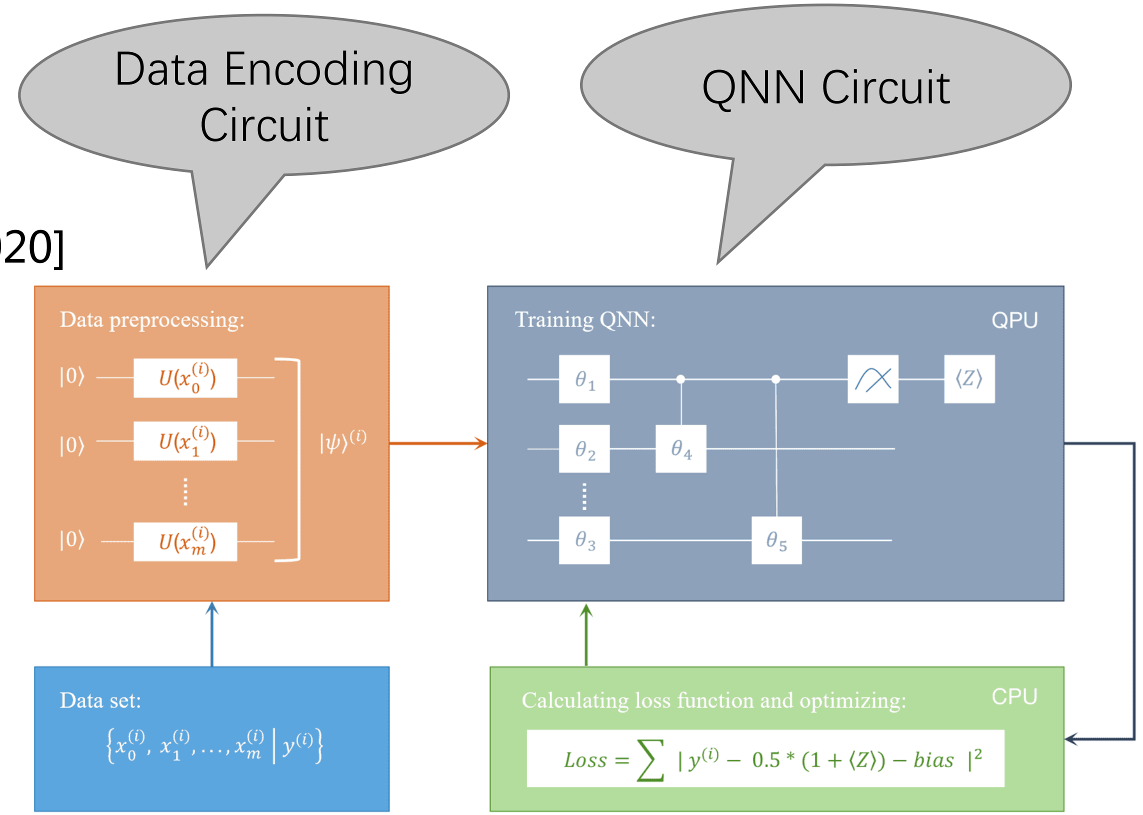
Motivation

- QNN architectures
- Data Encoding Circuit



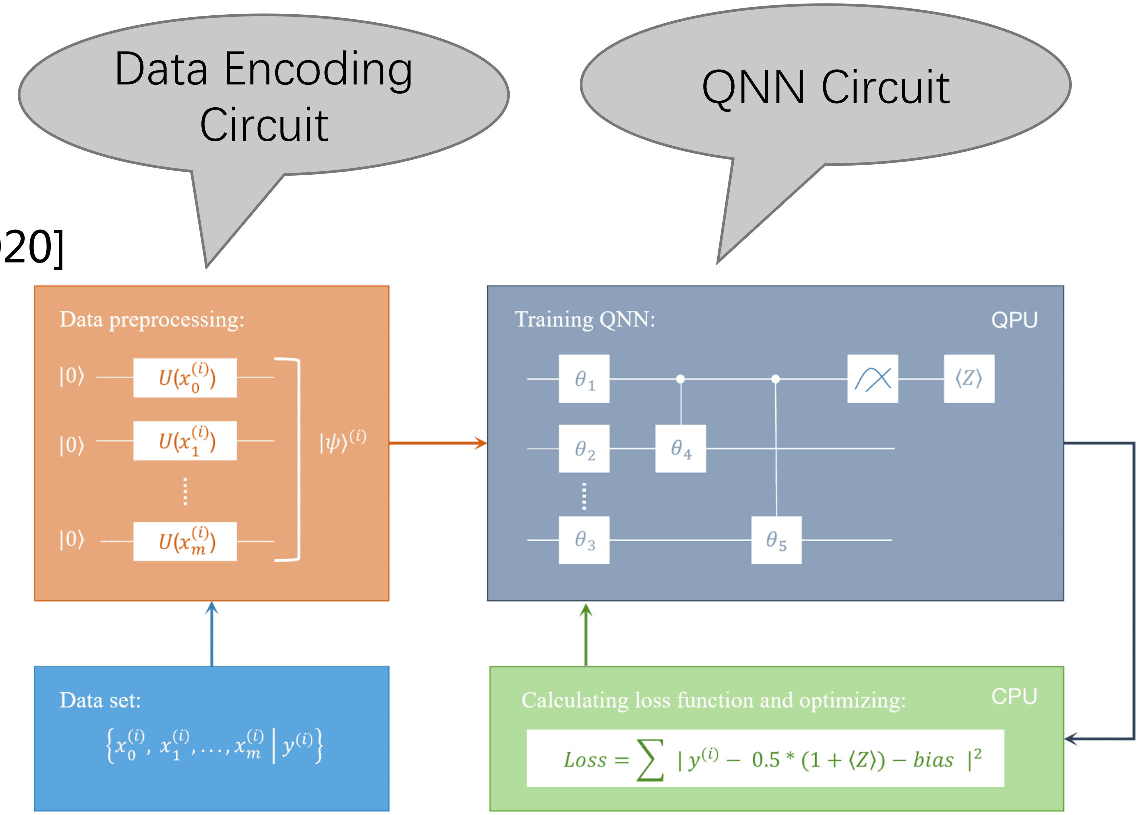
Motivation

- QNN architectures
  - Strongly Entangling Circuit [Schuld 2020]
  - QCNN [Cong et al 2019]
  - Tree-Tensor Network [Grant et al 2018]
  - Auto Search
- Data Encoding Circuit



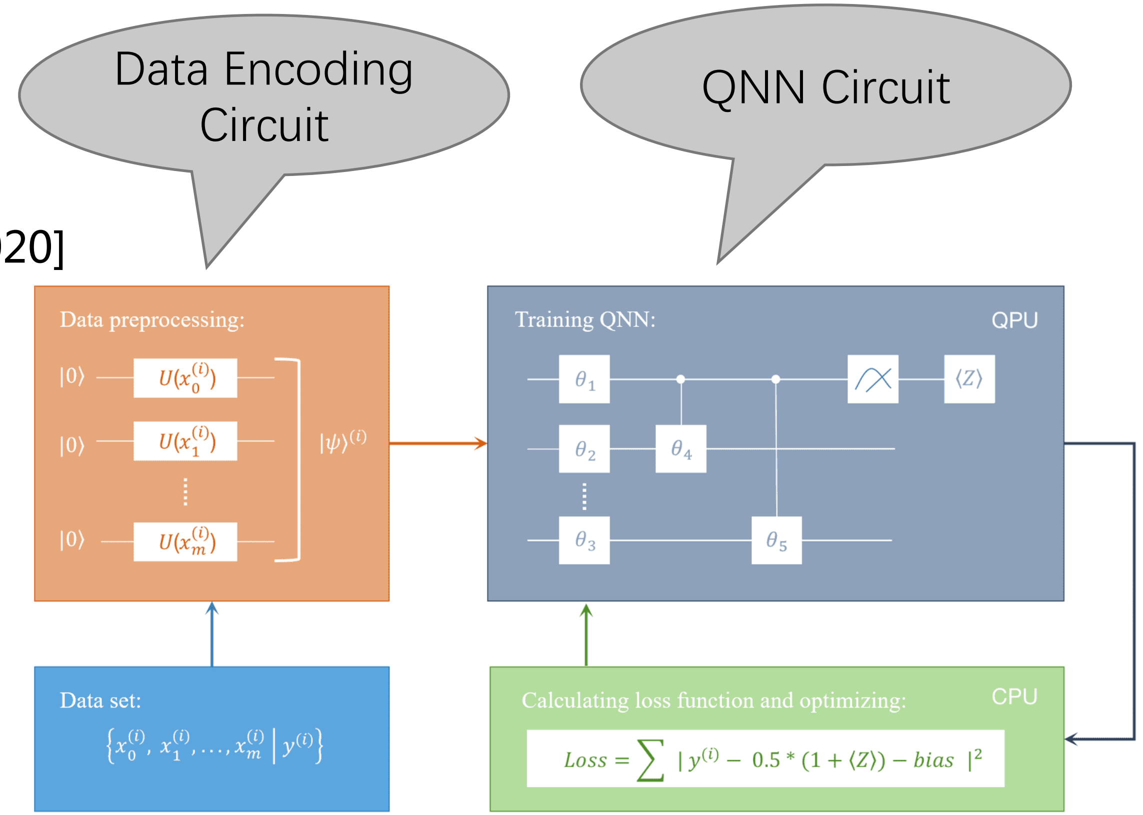
Motivation

- QNN architectures
  - Strongly Entangling Circuit [Schuld 2020]
  - QCNN [Cong et al 2019]
  - Tree-Tensor Network [Grant et al 2018]
  - Auto Search
- Data Encoding Circuit
  - Few literature



Motivation

- QNN architectures
  - Strongly Entangling Circuit [Schuld 2020]
  - QCNN [Cong et al 2019]
  - Tree-Tensor Network [Grant et al 2018]
  - Auto Search
- Data Encoding Circuit
  - Few literature
  - But it is also important
    - Kernel's perspective [Schuld 2021]
    - Influence the generalization error [Caro et al 2021, Banchi et al 2021]



This Work

- Focus
  - PQC-based data encoding strategy
- Question
  - How to systematically understand such encoding strategies?

Main results

- Concentration -- average encoded quantum state

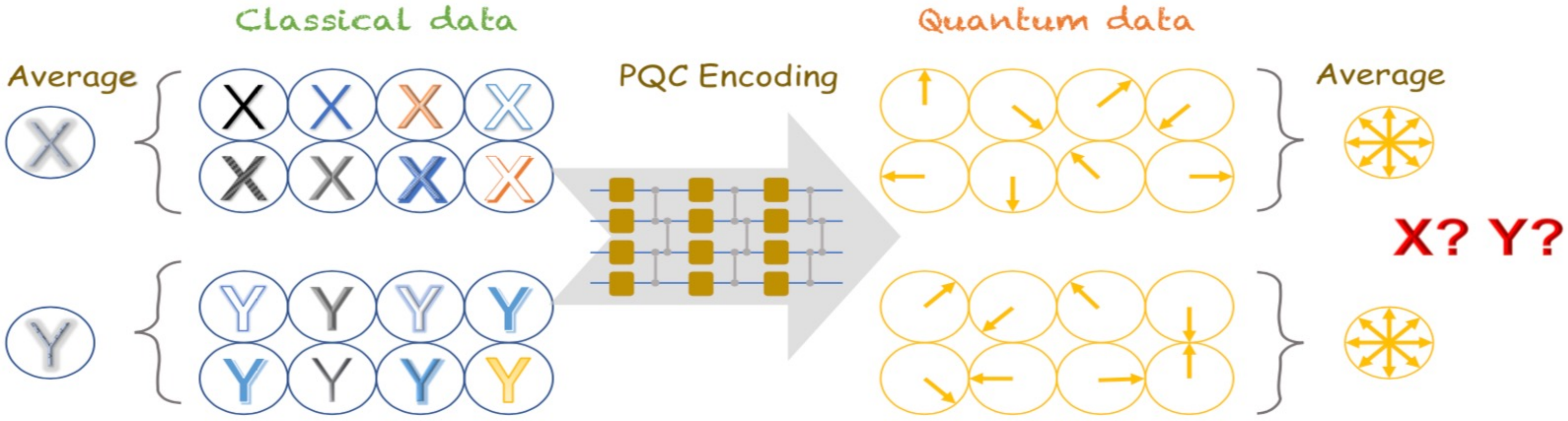
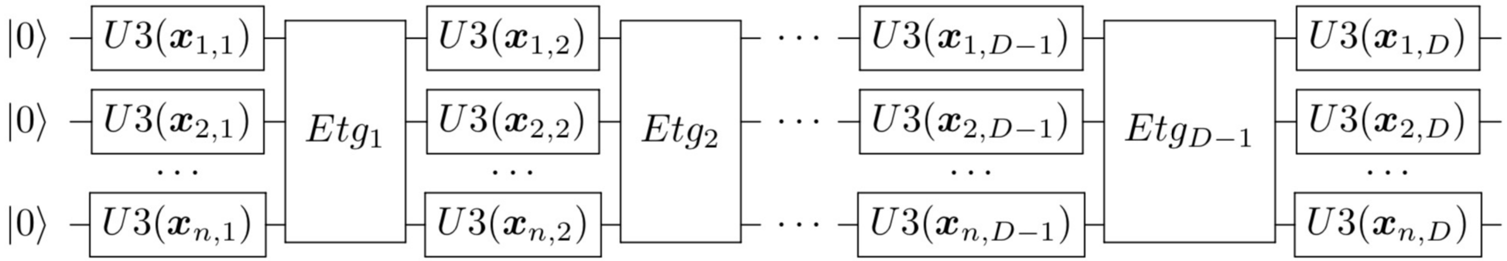


Figure 1: Cartoon illustrating the concentration of PQC-based data encoding. The average encoded quantum states concentrates on the maximally mixed state at an exponential rate on the encoding depth. This concentration implies the theoretical indistinguishability of the encoded quantum data.

Data Encoding Concentration

Theorem 2. ***(Data Encoding Concentration)*** *Assume each element of a $3nD$-dimensional vector $\boldsymbol{x}$ obeys an IGD, i.e., $x_{j,d,k} \sim \mathcal{N}(\mu_{j,d,k}, \sigma^2_{j,d,k})$, where $\sigma_{j,d,k} \geq \sigma$ for some constant σ and $1 \leq j \leq n, 1 \leq d \leq D, 1 \leq k \leq 3$. If $\boldsymbol{x}$ is encoded into an n-qubit pure state $\rho(\boldsymbol{x})$ according to the circuit in Fig. 3, the quantum divergence between the average encoded state $\bar{\rho}$ and the maximally mixed state $\mathbb{1}$ is upper-bounded as*

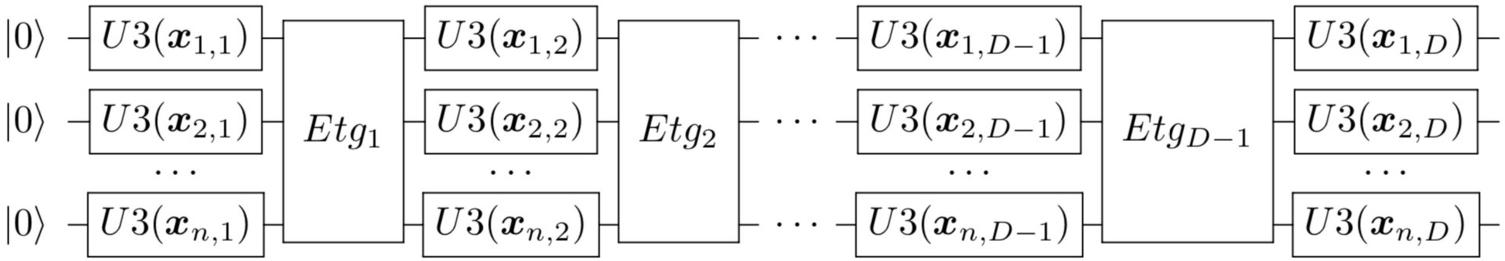
$$D_2\left(\bar{\rho} \| \mathbb{1}\right) \leq \log(1 + (2^n - 1)\mathrm{e}^{-D\sigma^2}). \tag{5}$$



Data Encoding Concentration

Theorem 2. ***(Data Encoding Concentration)*** *Assume each element of a $3nD$-dimensional vector $\boldsymbol{x}$ obeys an IGD, i.e., $x_{j,d,k} \sim \mathcal{N}(\mu_{j,d,k}, \sigma^2_{j,d,k})$, where $\sigma_{j,d,k} \geq \sigma$ for some constant σ and $1 \leq j \leq n, 1 \leq d \leq D, 1 \leq k \leq 3$. If $\boldsymbol{x}$ is encoded into an n-qubit pure state $\rho(\boldsymbol{x})$ according to the circuit in Fig. 3, the quantum divergence between the average encoded state $\bar{\rho}$ and the maximally mixed state $\mathbb{1}$ is upper-bounded as*

$$D_2\left(\bar{\rho} \| \mathbb{1}\right) \leq \log(1 + (2^n - 1)\mathrm{e}^{-D\sigma^2}). \tag{5}$$



- Quantum 2-relative Renyi divergence decays exponentially in depth *D*

$$D_2(\rho \| \sigma) = \log \operatorname{Tr}\left[\rho^2 \sigma^{-1}\right]$$

Applications in Quantum Supervised Learning

- In the Scenario of Quantum Classification

Applications in Quantum Supervised Learning

- In the Scenario of Quantum Classification
  - Loss function

$$L(\boldsymbol{\theta}; \mathcal{D}) \equiv \frac{1}{KM} \sum_{m=1}^{KM} L^{(m)} \quad \text{with} \quad L^{(m)}\left(\boldsymbol{\theta}; (\boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)})\right) \equiv -\sum_{k=1}^{K} y_k^{(m)} \ln \frac{\mathrm{e}^{h_k}}{\sum_{j=1}^{K} \mathrm{e}^{h_j}}, \qquad (8)$$

$$h_k\left(\boldsymbol{x}^{(m)}, \boldsymbol{\theta}\right) = \mathrm{Tr}\left[H_k U(\boldsymbol{\theta}) \rho(\boldsymbol{x}^{(m)}) U^{\dagger}(\boldsymbol{\theta})\right]$$

Proposition 4. *Consider a K-classification task with the data set $\mathcal{D}$ defined in Def. 3. If the encoding depth $D \geq \frac{1}{\sigma^2}[(n+4)\ln 2 + 2\ln(1/\epsilon)]$ for some $\epsilon \in (0,1)$, then the partial gradient of the loss function defined in Eq. (8) with respect to each parameter θ_i of the employed QNN is bounded as*

$$\left|\frac{\partial L(\boldsymbol{\theta}; \mathcal{D})}{\partial \theta_i}\right| \leq K\epsilon \qquad (10)$$

with a probability of at least $1 - 2\mathrm{e}^{-M\epsilon^2/8}$.

Applications in Quantum Supervised Learning

- In the Scenario of Quantum Classification
  - Loss function

$$L(\boldsymbol{\theta}; \mathcal{D}) \equiv \frac{1}{KM} \sum_{m=1}^{KM} L^{(m)} \quad \text{with} \quad L^{(m)}\left(\boldsymbol{\theta}; (\boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)})\right) \equiv -\sum_{k=1}^{K} y_k^{(m)} \ln \frac{\mathrm{e}^{h_k}}{\sum_{j=1}^{K} \mathrm{e}^{h_j}}, \qquad (8)$$

$$h_k\left(\boldsymbol{x}^{(m)}, \boldsymbol{\theta}\right) = \mathrm{Tr}\left[H_k U(\boldsymbol{\theta}) \rho(\boldsymbol{x}^{(m)}) U^\dagger(\boldsymbol{\theta})\right]$$

Proposition 4. *Consider a K-classification task with the data set $\mathcal{D}$ defined in Def. 3. If the encoding depth $D \geq \frac{1}{\sigma^2}\left[(n+4)\ln 2 + 2\ln(1/\epsilon)\right]$ for some $\epsilon \in (0,1)$, then the partial gradient of the loss function defined in Eq. (8) with respect to each parameter θ_i of the employed QNN is bounded as*

$$\left|\frac{\partial L(\boldsymbol{\theta}; \mathcal{D})}{\partial \theta_i}\right| \leq K\epsilon \qquad (10)$$

with a probability of at least $1 - 2\mathrm{e}^{-M\epsilon^2/8}$.

- Limit the trainability (probably classification ability) of QNNs

Applications in Quantum Supervised Learning

- In the Scenario of Quantum State Discrimination

Applications in Quantum Supervised Learning

- In the Scenario of Quantum State Discrimination
  - Success probability

$$p_{\text{succ}} \equiv \max_{\{\Pi_k\}_k} \frac{1}{K} \sum_{k=1}^{K} \mathrm{Tr}\left[\Pi_k \bar{\rho}_{k,M}\right] \quad \text{with} \quad \bar{\rho}_{k,M} \equiv \frac{1}{M} \sum_{m=1}^{KM} y_k^{(m)} \rho(\boldsymbol{x}^{(m)}),$$

Proposition 5. *Consider a K-class discrimination task with the data set $\mathcal{D}$ defined in Def. 3. If the encoding depth $D \geq \frac{1}{\sigma^2}\left[(n+4)\ln 2 + 2\ln(1/\epsilon)\right]$ for a given $\epsilon \in (0,1)$, then with a probability of at least $1 - 2\mathrm{e}^{-M\epsilon^2/8}$, the maximum success probability p_{succ} is bounded as*

$$p_{\text{succ}} \leq 1/K + \epsilon. \tag{12}$$

Applications in Quantum Supervised Learning

- In the Scenario of Quantum State Discrimination
  - Success probability

$$p_{\text{succ}} \equiv \max_{\{\Pi_k\}_k} \frac{1}{K} \sum_{k=1}^{K} \mathrm{Tr}\left[\Pi_k \bar{\rho}_{k,M}\right] \quad \text{with} \quad \bar{\rho}_{k,M} \equiv \frac{1}{M} \sum_{m=1}^{KM} y_k^{(m)} \rho(\boldsymbol{x}^{(m)}),$$

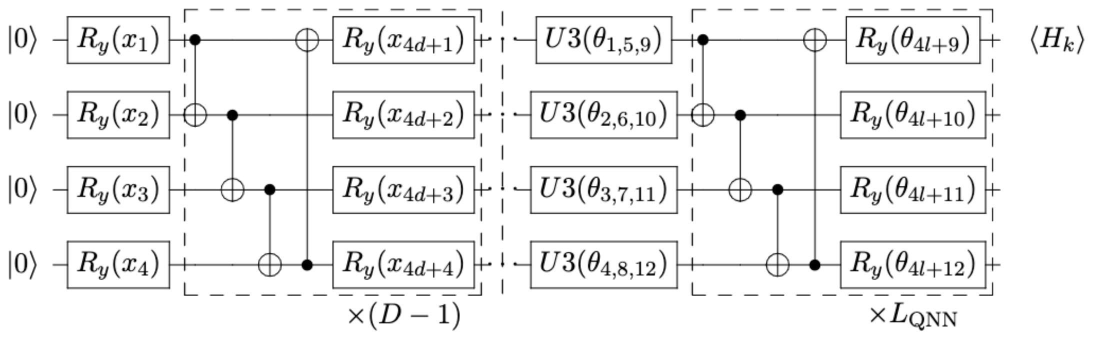
Proposition 5. *Consider a K-class discrimination task with the data set $\mathcal{D}$ defined in Def. 3. If the encoding depth $D \geq \frac{1}{\sigma^2}\left[(n+4)\ln 2 + 2\ln(1/\epsilon)\right]$ for a given $\epsilon \in (0,1)$, then with a probability of at least $1 - 2\mathrm{e}^{-M\epsilon^2/8}$, the maximum success probability p_{succ} is bounded as*

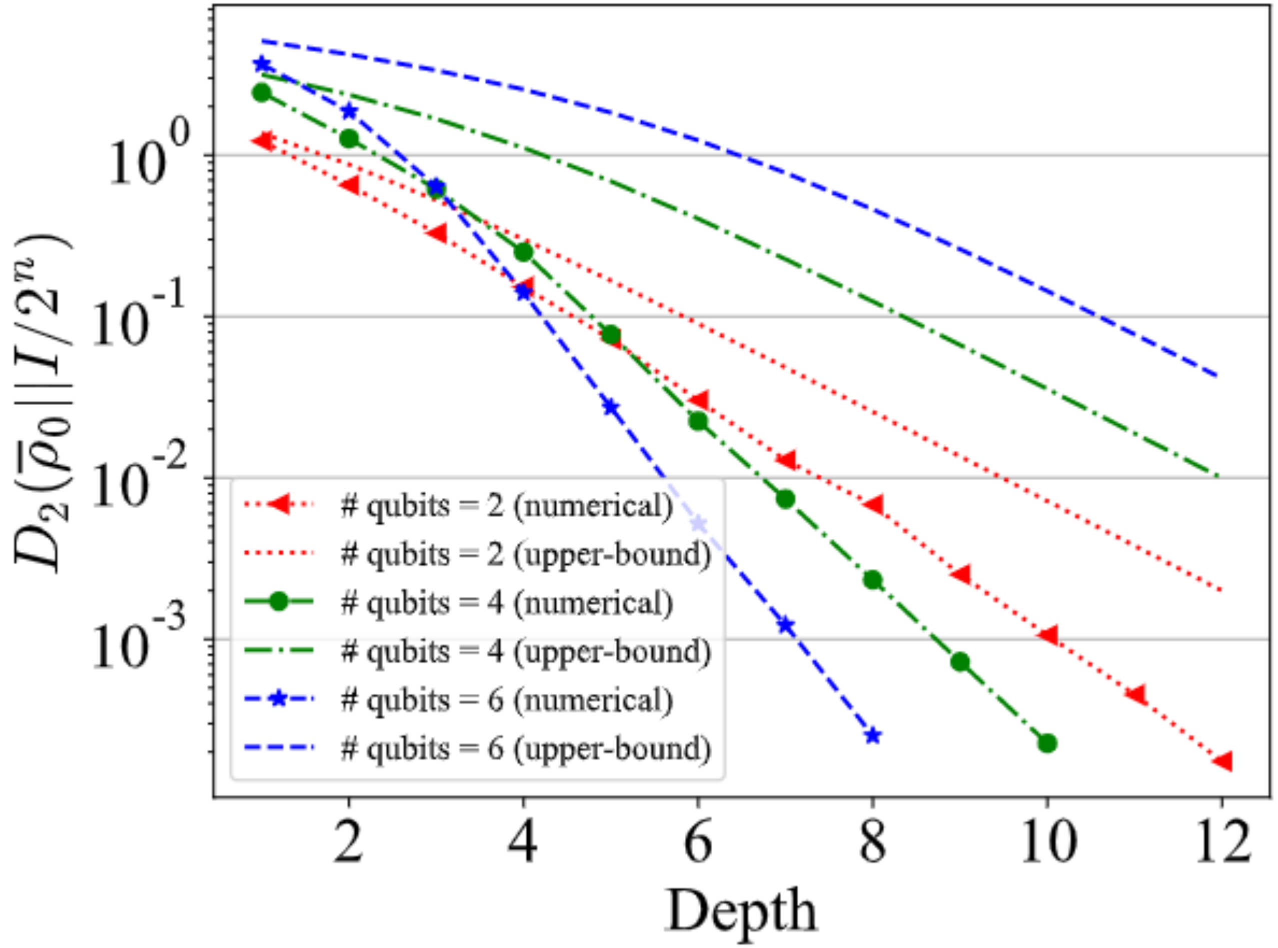
$$p_{\text{succ}} \leq 1/K + \epsilon. \tag{12}$$

- Restrict the distinguishability of POVMs

Numerical Experiments (1)

- Numerical quantum divergences indeed decrease exponentially for the following PQCs





Numerical Experiments (2)

- Synthetic Data Set – Test accuracy

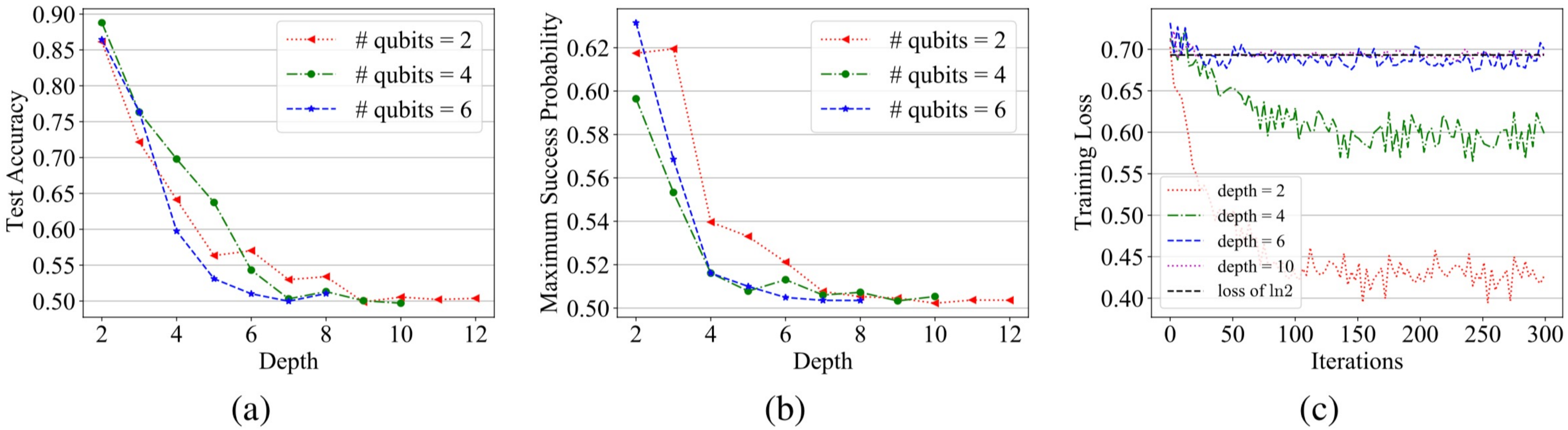


Figure 6: Numerical results for synthetic data sets under the encoding strategy in Fig. 4. In all qubit cases, (a) the test accuracy of QNN (or (b) the maximum success probability of POVM) will eventually decay to 50% or so with the depth growing; (c) In the 4 qubit case, for instance, the training losses of QNN do not decrease and stay at about $\ln 2$ in the training process when the depth becomes large enough.

Numerical Experiments (2)

- Synthetic Data Set – Test accuracy

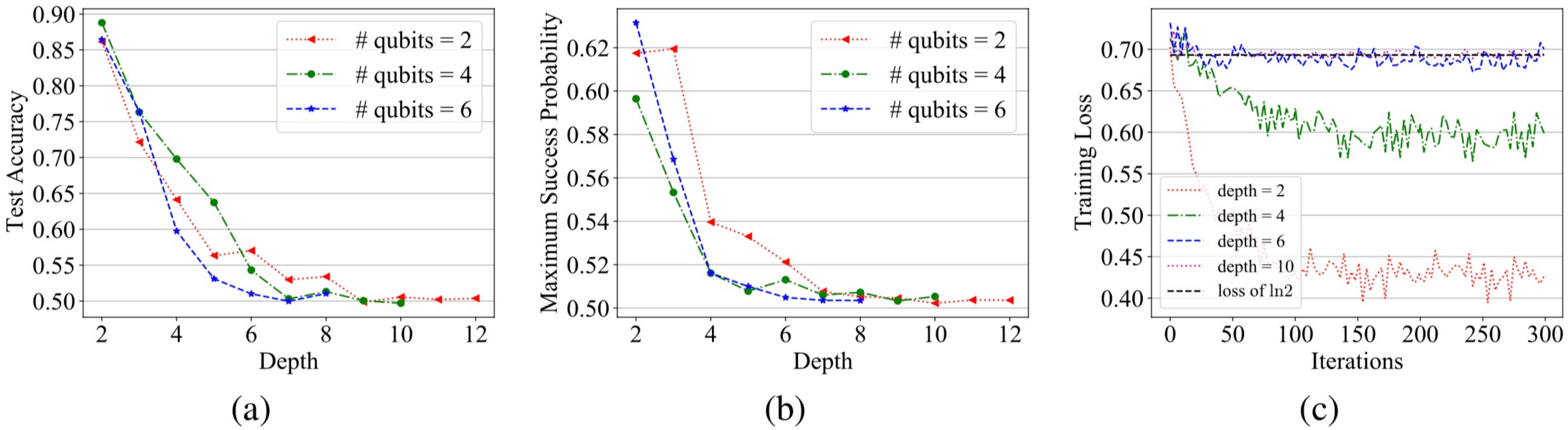


Figure 6: Numerical results for synthetic data sets under the encoding strategy in Fig. 4. In all qubit cases, (a) the test accuracy of QNN (or (b) the maximum success probability of POVM) will eventually decay to 50% or so with the depth growing; (c) In the 4 qubit case, for instance, the training losses of QNN do not decrease and stay at about $\ln 2$ in the training process when the depth becomes large enough.

- These results are in line with our theoretical analysis, i.e., limitations of PQCs

Numerical Experiments (3)

- Public Data Set

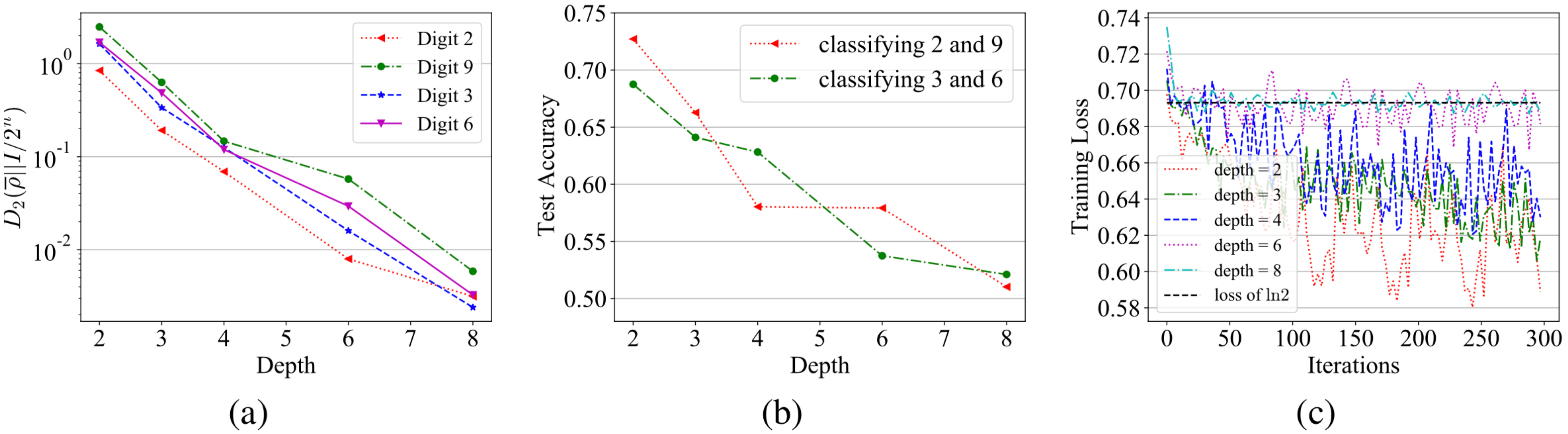


Figure 7: Numerical results of QNN for MNIST data set under the encoding strategy in Fig. 4. (a) The curves for the quantum divergence between the averaged encoded state $\bar{\rho}$ of each handwritten digit and the maximally mixed state $\mathbb{1}$ decrease exponentially on depth. (b) The test accuracy reduce rapidly with a larger encoding depth; (c) In the case of classifying digits 3 and 6, when the depth is large (e.g., 8), it is difficult to keep the training loss away from $\ln 2$ in the training process.

Conclusion

- This work explores the data encoding concentration by proving the exponential decay (in encoding depth) of the upper bound of quantum divergence.
- The quantum states encoded by deep PQCs will seriously limit the classification performance of downstream supervised learning tasks.
- This work also provides insights in developing nontrivial quantum encoding strategies, i.e., avoiding concentration.

Thank you!