QTML2023

Understanding QML also requires rethinking generalization

Elies Gil-Fuster, FU Berlin

@EliesMiquel

1. Quantum Machine Learning

Quantum model

+

Quantum data

+

Classical training algorithm

2. Generalization

Dictionary

Definitions from Oxford Languages · Learn more

generalization

/ˌdʒɛn(ə)rəlʌɪˈzeɪʃn/

noun

a general statement or concept obtained by inference from specific cases.
"he was making sweeping generalizations"

- the action of generalizing.
  "such anecdotes cannot be a basis for generalization"

Train with few data → learn patterns → try on new data

… → Succeed on new data!

3. Why/how generalization?

Train with few data → learn patterns → <u>Succeed</u> on new data!

… → why?

<u>Really difficult:</u>

Why/How do (Q)ML models generalize?

4. How? Model capacity!

Why/How do (Q)ML models generalize?

1st attempt:

It has to do with model capacity.

"Expressivity"

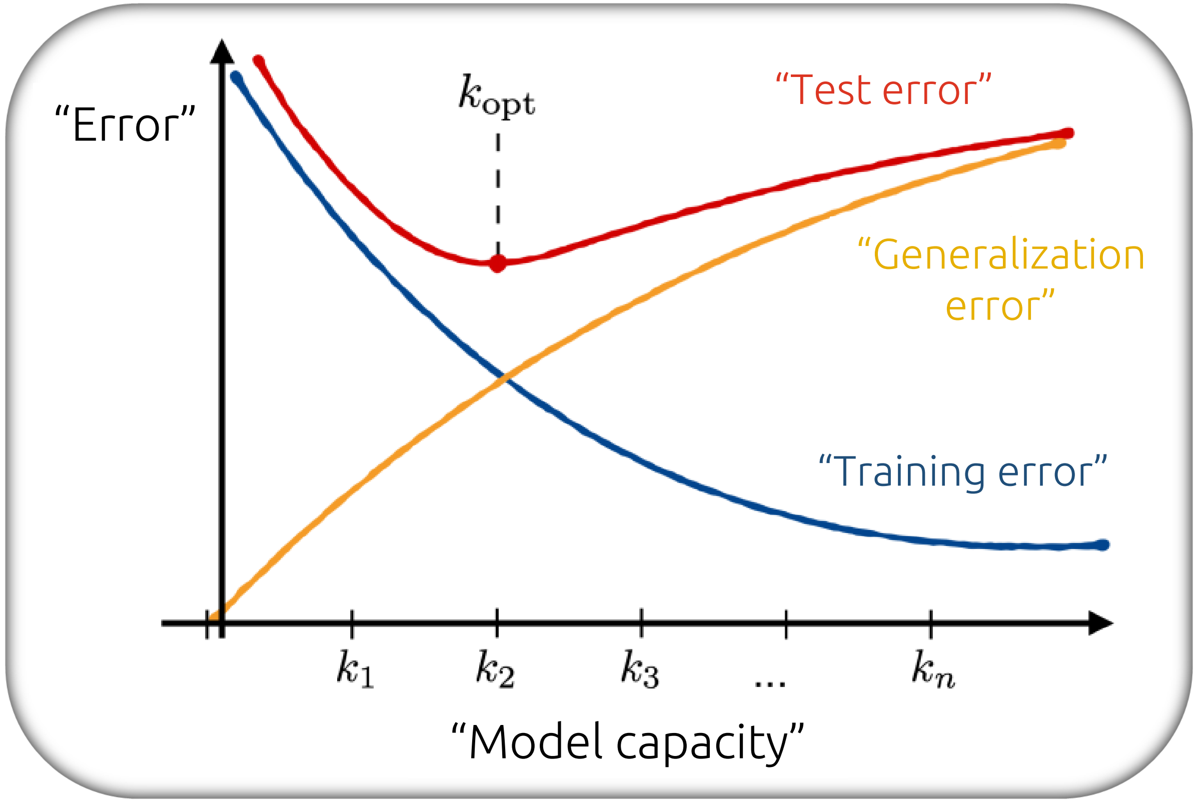
"Complexity measure"

"Models should be large enough, but not larger"

5. The traditional picture

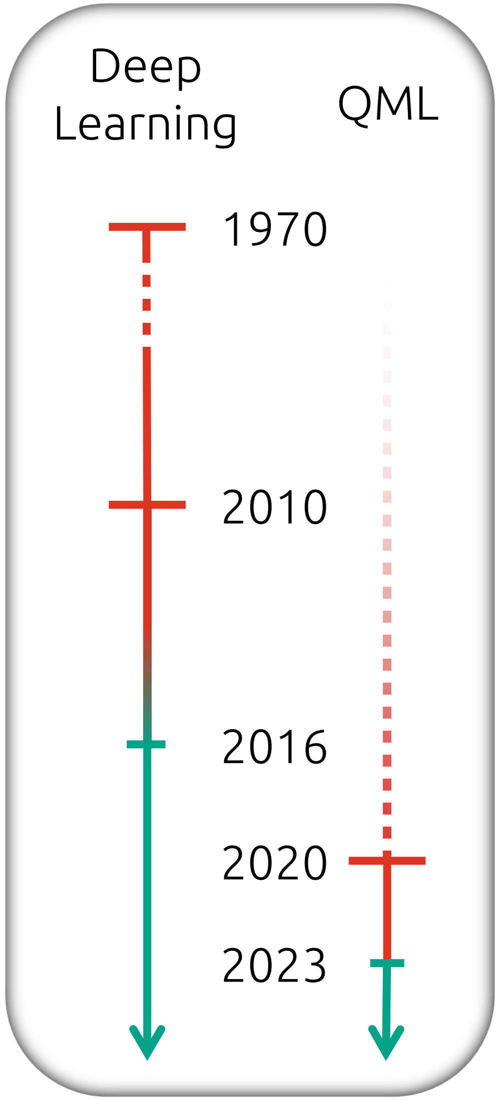
Underfitting
"Model capacity is not large enough"

Overfitting
"Model capacity is too large"



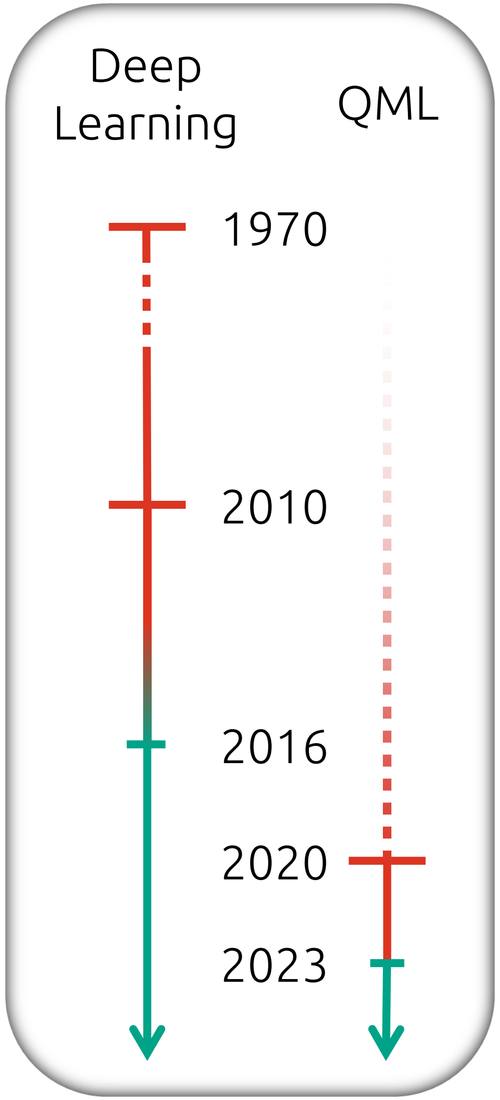
Caro et al. Quantum **5**, 582 (2021)

6. Context



1970s – 2010s: Traditional learning theory: model capacity.

6. Context

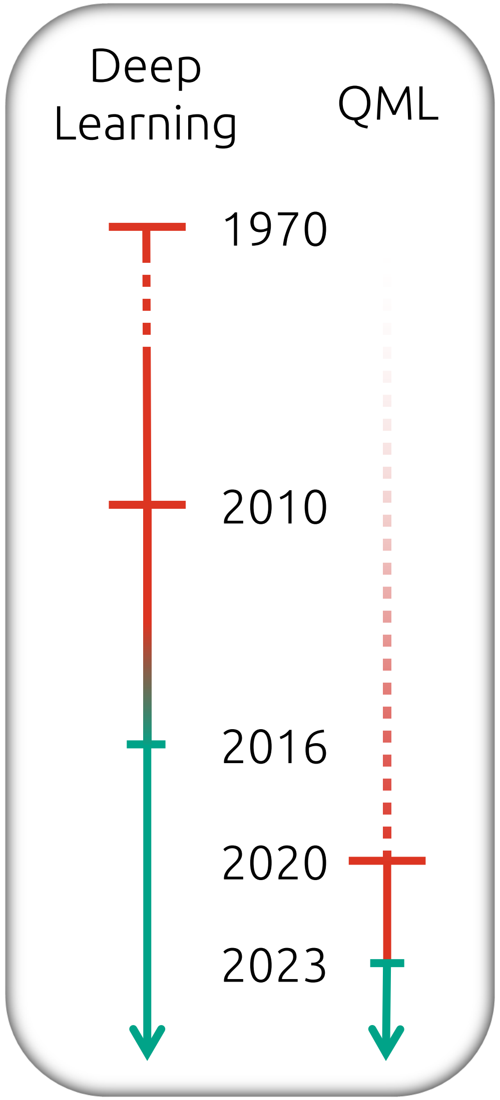


1970s – 2010s: Traditional learning theory: model capacity.

Zhang et al. (2016):

"Model capacity cannot explain generalization in Deep Learning".

6. Context



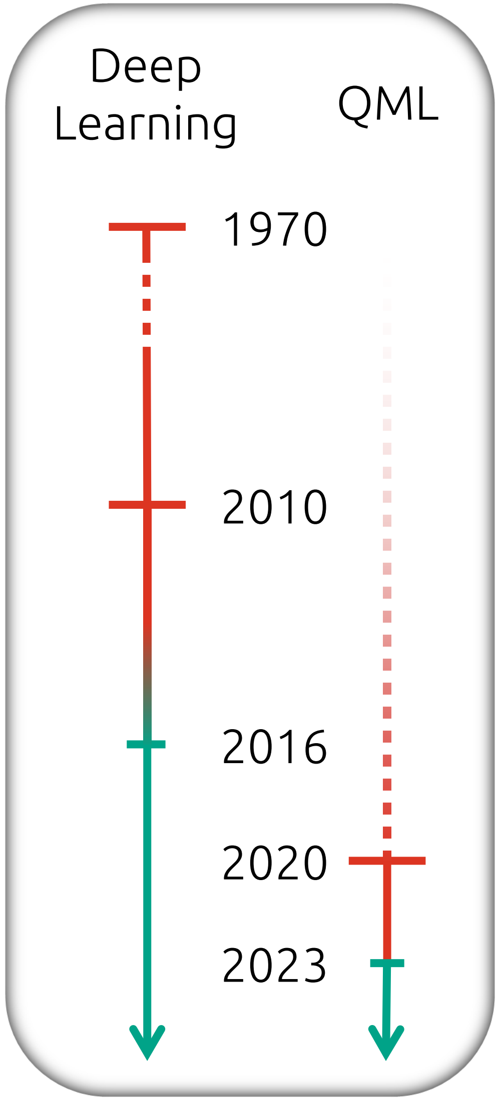
1970s – 2010s: Traditional learning theory: model capacity.

Zhang et al. (2016):

"Model capacity cannot explain generalization in Deep Learning".

2016 - ...: Modern approaches to generalization in classical ML.

6. Context



1970s – 2010s: Traditional learning theory: model capacity.

Zhang et al. (2016):

"Model capacity cannot explain generalization in Deep Learning".

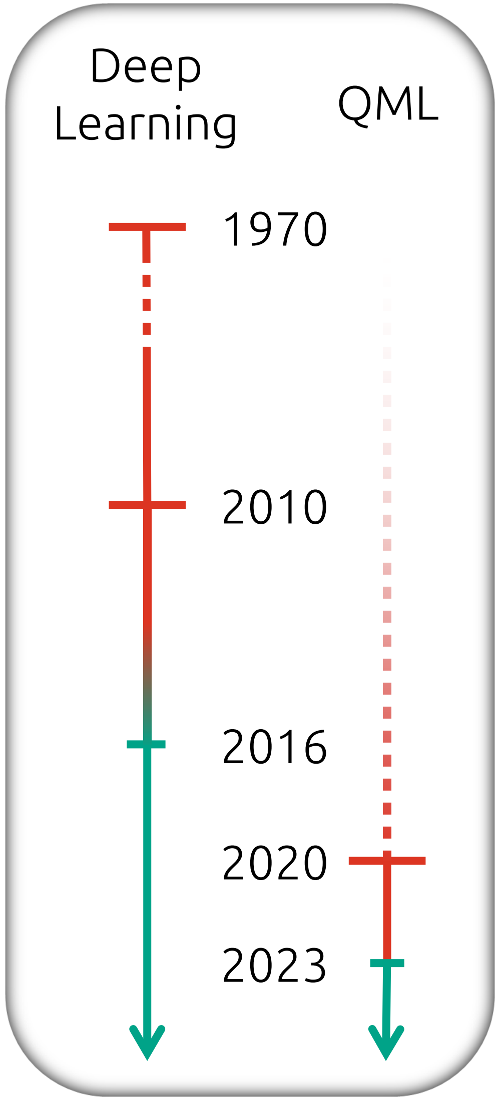
2016 - ...: Modern approaches to generalization in classical ML.

2020s: Traditional learning theory applied to QML.

Elephant:
Will it work?

Different models...
maybe?!

6. Context



1970s – 2010s: Traditional learning theory: model capacity.

Zhang et al. (2016):

"Model capacity cannot explain generalization in Deep Learning".

2016 - ...: Modern approaches to generalization in classical ML.

2020s: Traditional learning theory applied to QML.

Elephant:
Will it work?

Different models...
maybe?!

Our work (2023):

"Model capacity cannot explain generalization in current QML".

7. Our contribution

Hypothesis:

Generalization has to do with model capacity.

Test:

“Does model capacity explain generalization?”

Results:

“Model capacity cannot explain generalization in current QML”.

8. Experiment

Results: “*Model capacity cannot explain generalization in current QML*”.

How? What if input data and labels are uncorrelated?

8. Experiment

Results: "*Model capacity cannot explain generalization in current QML*".

How? What if input data and labels are uncorrelated?

Model can fit original training sets

$\Downarrow$

good generalization (?)

8. Experiment

Results: "*Model capacity <u>cannot explain</u> generalization in current QML*".

How? What if input data and labels are uncorrelated?

Model can fit original training sets $\Downarrow$ good generalization (?)	Model can fit corrupt training sets $\Downarrow$ bad generalization

8. Experiment

Results: "*Model capacity cannot explain generalization in current QML*".

How? What if input data and labels are uncorrelated?

Model can fit original training sets $\Downarrow$ good generalization (?)	Model can fit corrupt training sets $\Downarrow$ bad generalization

Model can do both $\Rightarrow$ model capacity indep of generalization

9. Take-home message

We still don't know why/how QML models generalize.

(So far QML literature *only* model capacity approach)

Invitation: Let's go beyond model capacity!

- Properties of the loss landscape
- Stability/robustness of optimization algorithm
- Shape of the learning curve
- …

References

Our work, “model capacity fails for current QML”:

Gil-Fuster, Eisert, Bravo-Prieto *arXiv:2306.13461 (2023)*

“Model capacity fails for deep NNs”:

Zhang et al. *arXiv:1611.03530 (2016)*

“PQCs have good generalization”:

Caro et al. *Nat. Comms.* ***13****, 4919 (2022)*

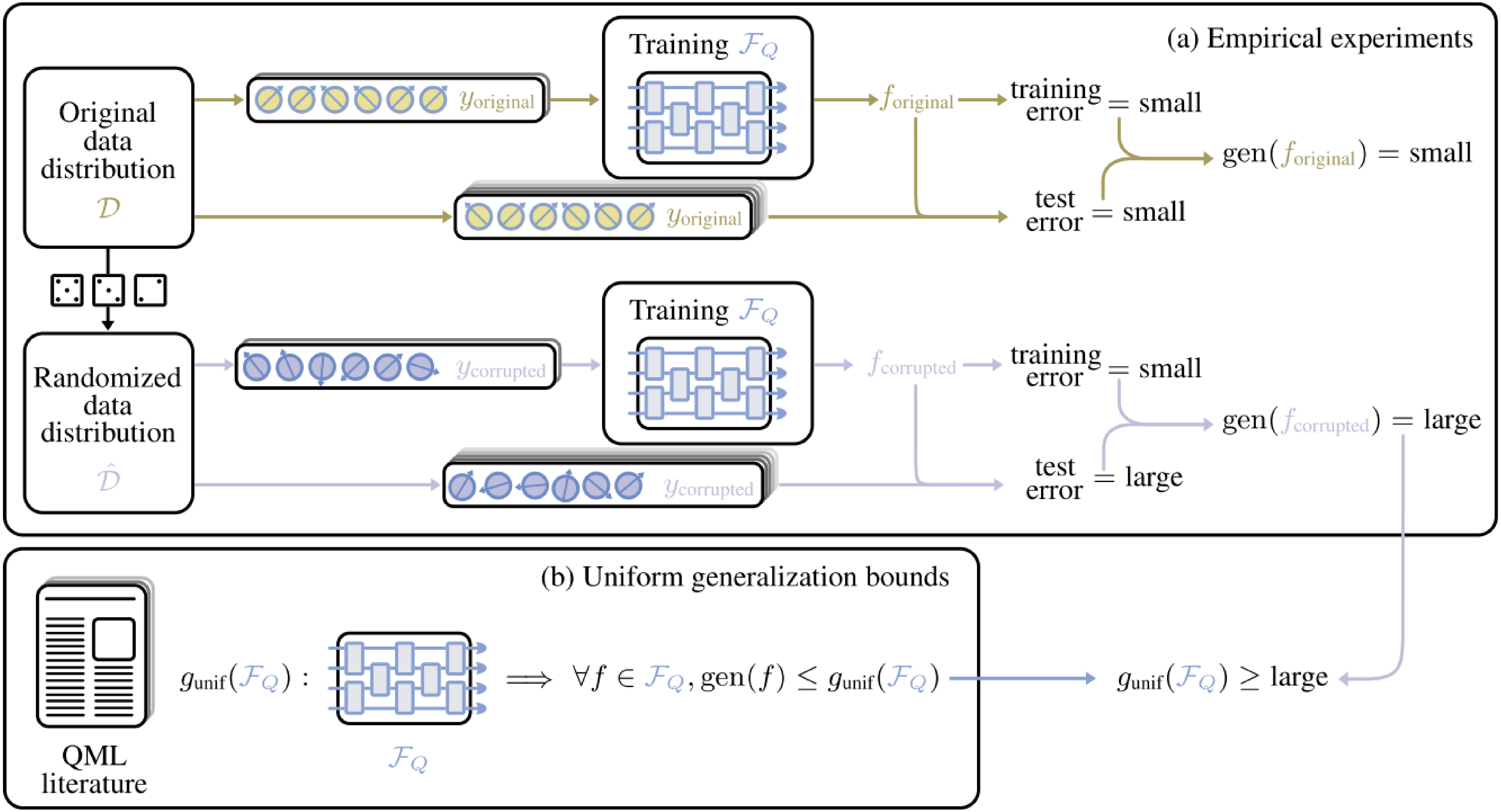
Thank you for your attention!

Understanding QML also requires rethinking generalization

arXiv:2306.13461

eliesgilfuster.eu

@EliesMiquel



(a) Empirical experiments
Original data distribution $\mathcal{D}$
y_{original}
Training $\mathcal{F}_Q$
f_{original}
training error = small
test error = small
$\text{gen}(f_{\text{original}}) = \text{small}$
Randomized data distribution $\hat{\mathcal{D}}$
$y_{\text{corrupted}}$
$f_{\text{corrupted}}$
training error = small
test error = large
$\text{gen}(f_{\text{corrupted}}) = \text{large}$
(b) Uniform generalization bounds
QML literature
$g_{\text{unif}}(\mathcal{F}_Q)$:
$\mathcal{F}_Q$
$\Longrightarrow \forall f \in \mathcal{F}_Q, \text{gen}(f) \leq g_{\text{unif}}(\mathcal{F}_Q)$
$g_{\text{unif}}(\mathcal{F}_Q) \geq \text{large}$

[2306.13461]

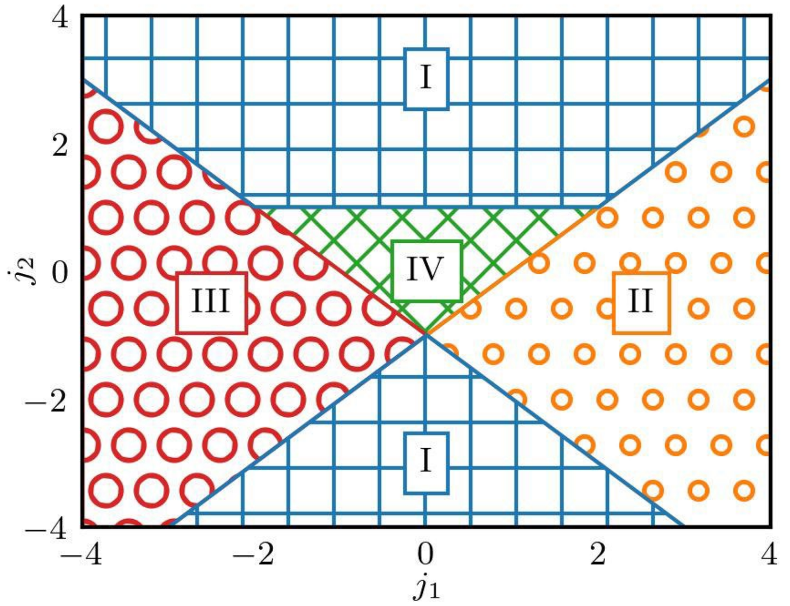
Quantum phase recognition with QCNN

Learning task: classify ground states of Generalized cluster Hamiltonian according to their phase.

$$H(j_1, j_2) = \sum_{j=1}^{n} \left(Z_j - j_1 X_j X_{j+1} + j_2 X_{j-1} Z_j X_{j+1} \right).$$

Ground states $\rho(j_1, j_2)$ four phases:

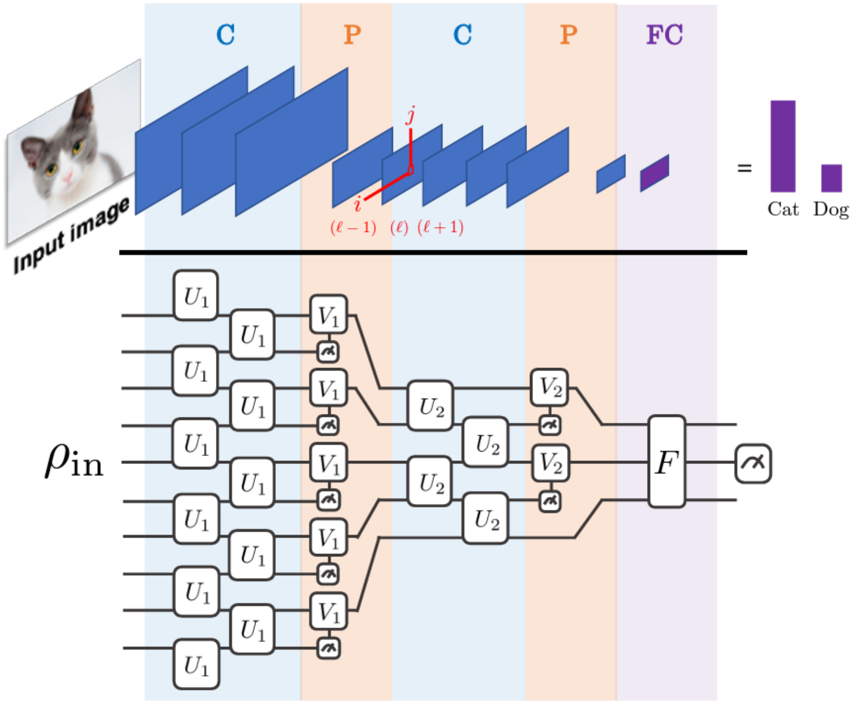
I. Symmetry-protected topological
II. Ferromagnetic
III. Anti-ferromagnetic
IV. Trivial



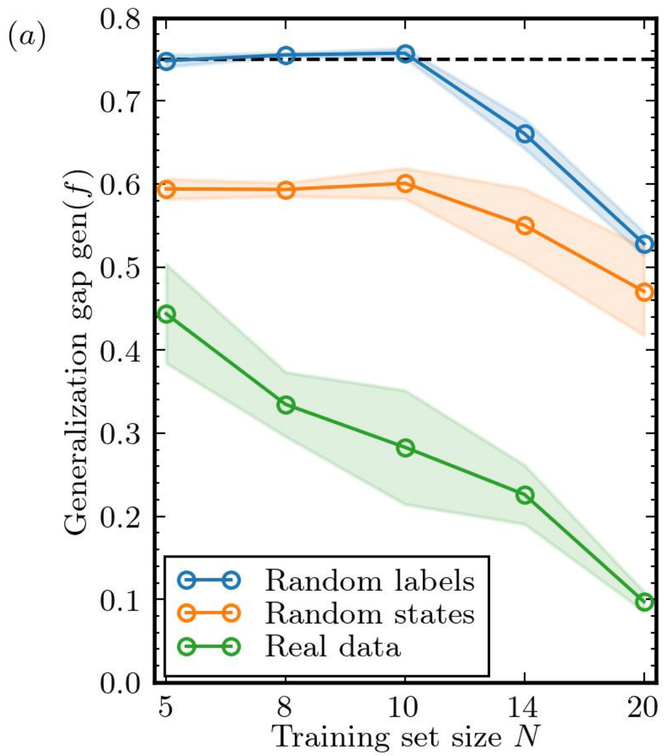
Quantum phase recognition with QCNN

Quantum Convolutional Neural Network (QCNN).

- Our quantum model $\mathcal{F}_Q$.
- Inspired in classical Deep Learning.
- Convolutional and pooling layers.
- Parameter sharing.
- Translation invariance.
- Exploit local correlations.
- No Barren Plateaus.



Numerical experiments



Average number of qubits $n \in \{8, 16, 32\}$.

Random guessing balanced classes

$$\mathrm{gen}(f) \approx 0.75$$

Random guessing unbalanced classes

$$\mathrm{gen}(f) \approx 0.6$$

Message: QCNN can fit random distributions

Memorization

Theorem (*Finite sample expressivity of PQCs– informal*).

Let $\rho_1, \dots, \rho_N$ be unknown quantum states on $n \in \mathbb{N}$ qubits, $N \in \mathcal{O}\big(\mathrm{poly}(n)\big)$, and fulfilling a distinguishability condition.

Then, we can construct a PQC of depth $\mathrm{poly}(n)$ as a parametrized observable $\mathcal{M}(\vartheta)$ such that,

for any $y_1, \dots, y_N \in \mathbb{R}$ real labels, we can find parameters ϑ_y fulfilling

$$\mathrm{tr}\left(\rho_i \mathcal{M}\left(\vartheta_y\right)\right) = y_i.$$