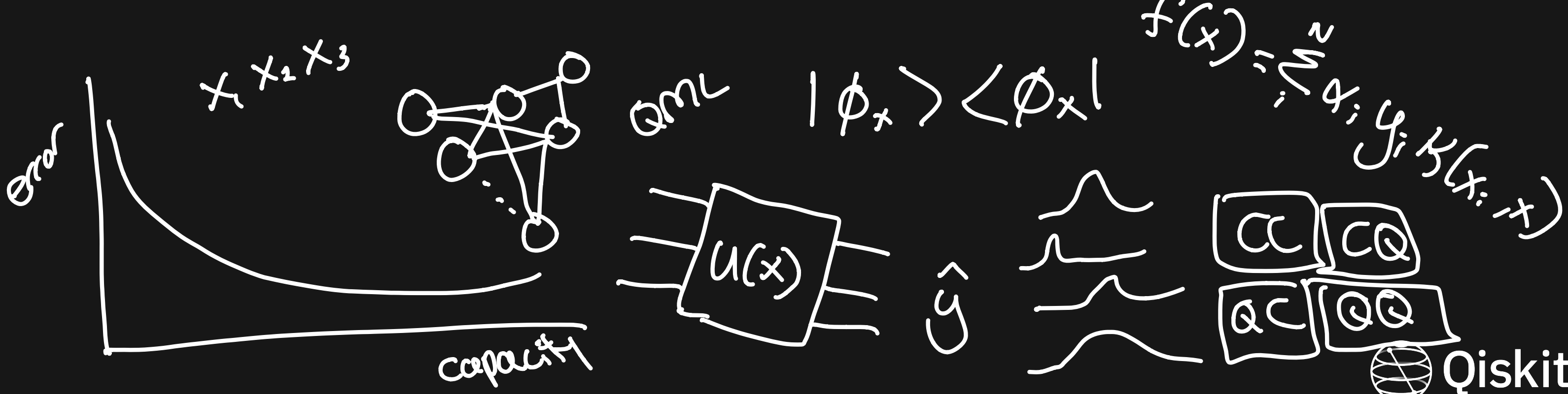
The capacity and power of quantum machine learning

Amira Abbas

IBM Quantum, University of KwaZulu-Natal



Qiskit

What does capacity mean?

In classical machine learning, capacity is used interchangeably with many different terms

What does capacity mean?

In classical machine learning, capacity is used interchangeably with many different terms

Statistical complexity

Expressivity

Power

What does capacity mean?

In classical machine learning, capacity is used interchangeably with many different terms

Statistical complexity

Expressivity

Power

How many functions can my model approximate?

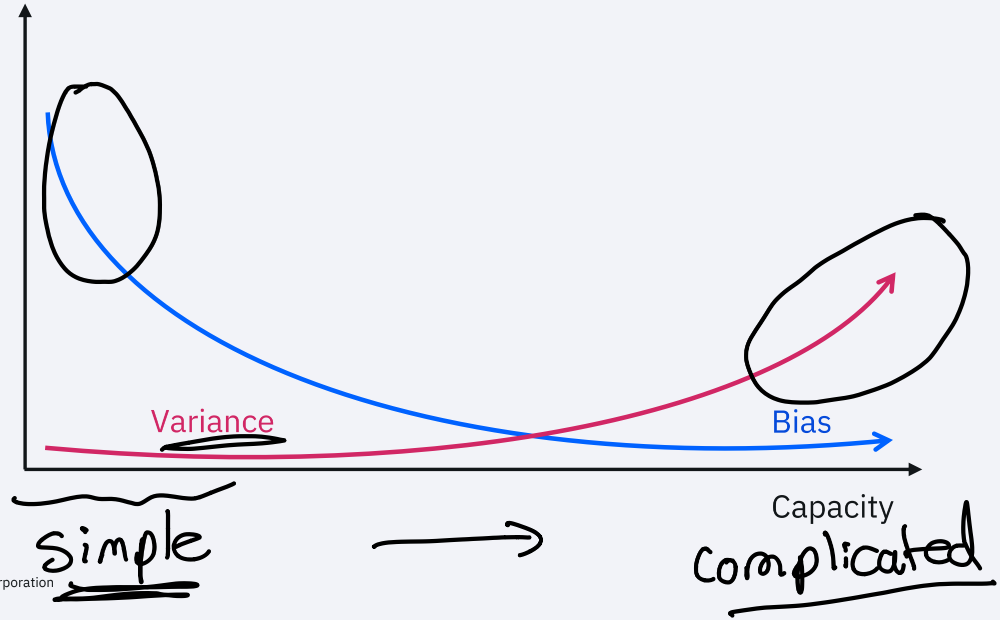
Is a higher capacity always better?

Generalization

We can understand capacity in the context of generalization performance

Generalization

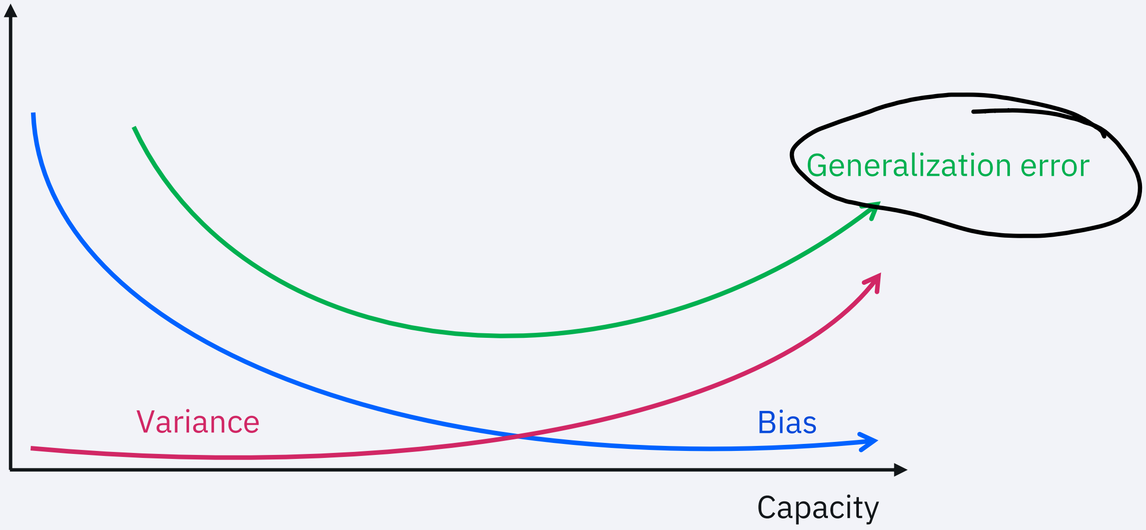
We can understand capacity in the context of generalization performance



Generalization

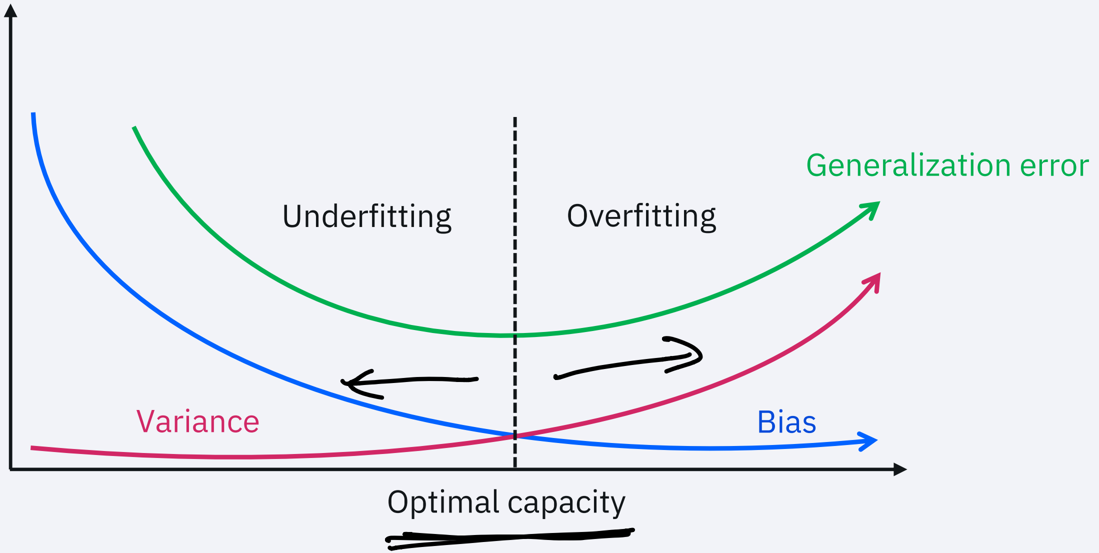
Qiskit

We can understand capacity in the context of generalization performance



Generalization

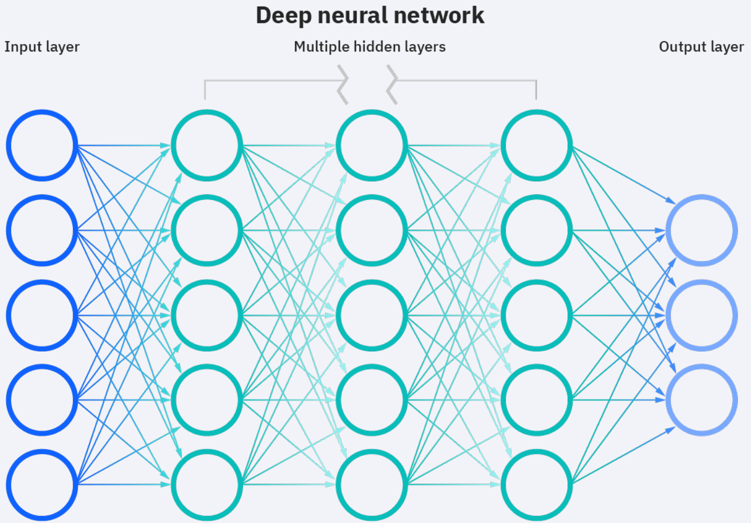
We can understand capacity in the context of generalization performance



How do we measure capacity? (Classically)

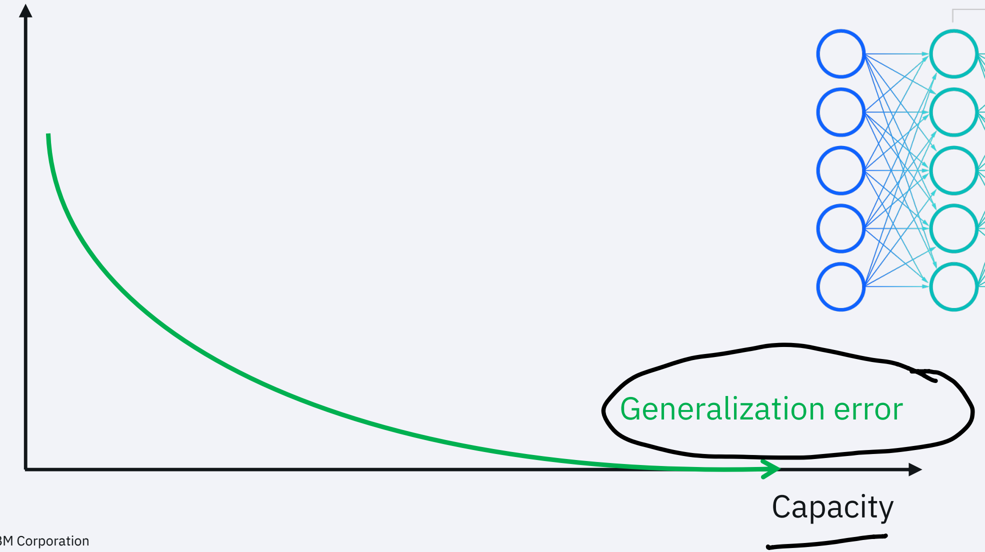
Capacity measures

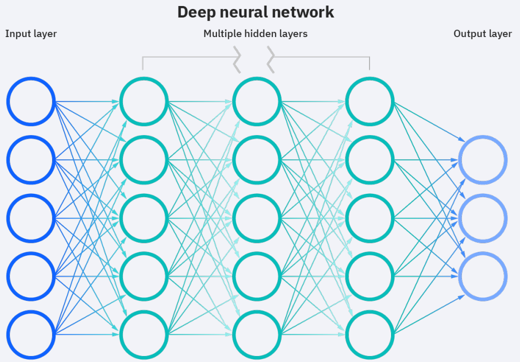
As a start, making your model larger (adding more parameters, increasing the depth) was assumed to add to capacity



Deep neural networks

A common misconception or seemingly paradoxical behaviour was that DNNs do not overfit





Deep neural networks

An important experiment shown by Zhang et al. showed that DNNs do overfit if trained on *random labels*

Understanding deep learning requires rethinking generalization

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, Oriol Vinyals

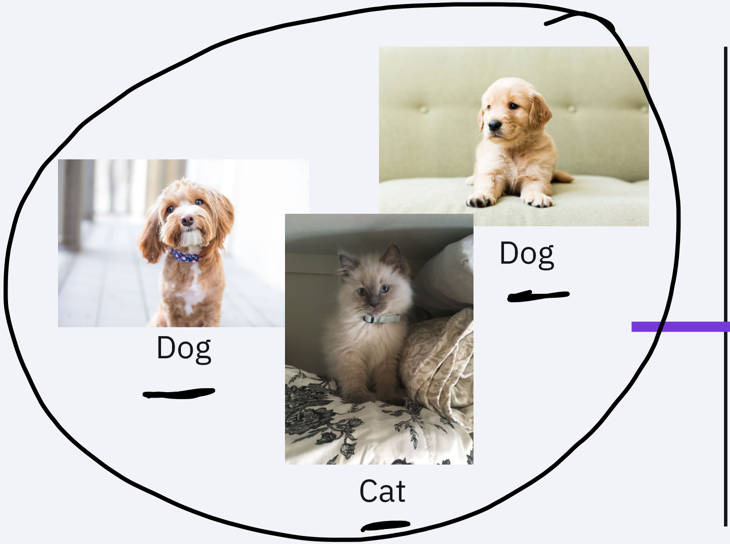
Despite their massive size, successful deep artificial neural networks can exhibit a remarkably small difference between training and test performance. Conventional wisdom attributes small generalization error either to properties of the model family, or to the regularization techniques used during training.
Through extensive systematic experiments, we show how these traditional approaches fail to explain why large neural networks generalize well in practice. Specifically, our experiments establish that state-of-the-art convolutional networks for image classification trained with stochastic gradient methods easily fit a random labeling of the training data. This phenomenon is qualitatively unaffected by explicit regularization, and occurs even if we replace the true images by completely unstructured random noise. We corroborate these experimental findings with a theoretical construction showing that simple depth two neural networks already have perfect finite sample expressivity as soon as the number of parameters exceeds the number of data points as it usually does in practice.
We interpret our experimental findings by comparison with traditional models.

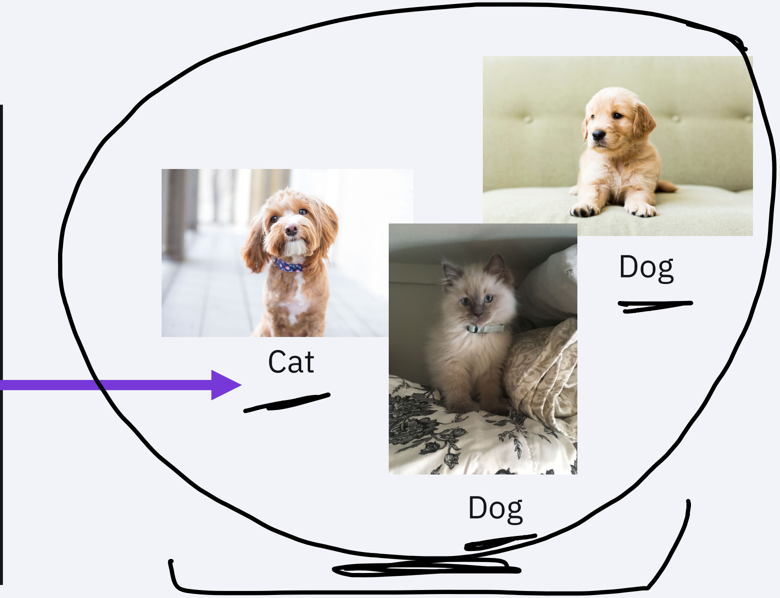
Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530.

Deep neural networks

Qiskit

An important experiment shown by Zhang et al. showed that DNNs do overfit if trained on *random labels*

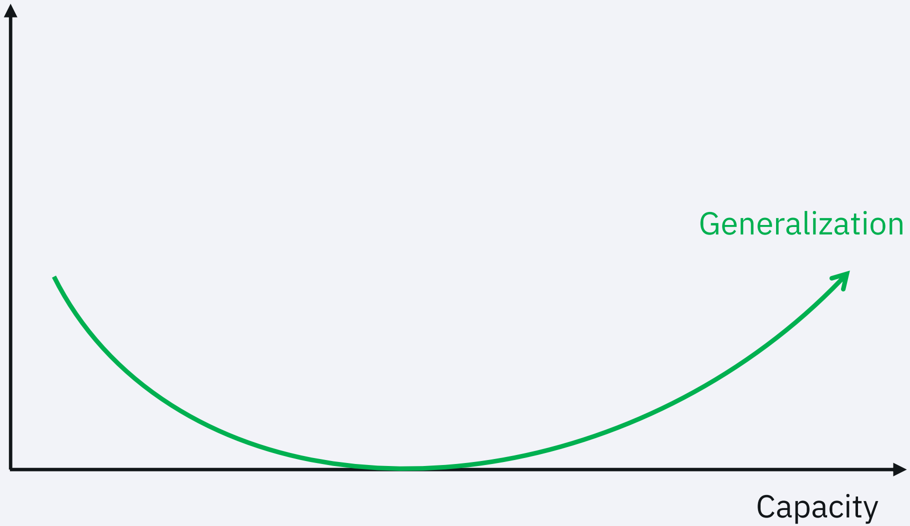




Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530.

Deep neural networks

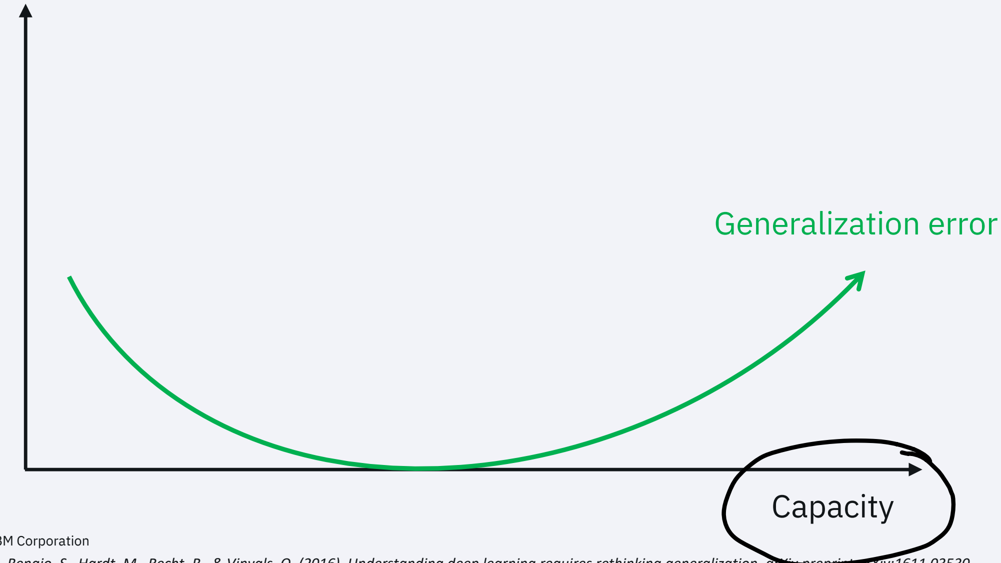
An important experiment shown by Zhang et al. showed that DNNs do overfit if trained on *random labels*



Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530.

Deep neural networks

This suggests that we have been thinking about capacity incorrectly and capacity measures must somehow incorporate data

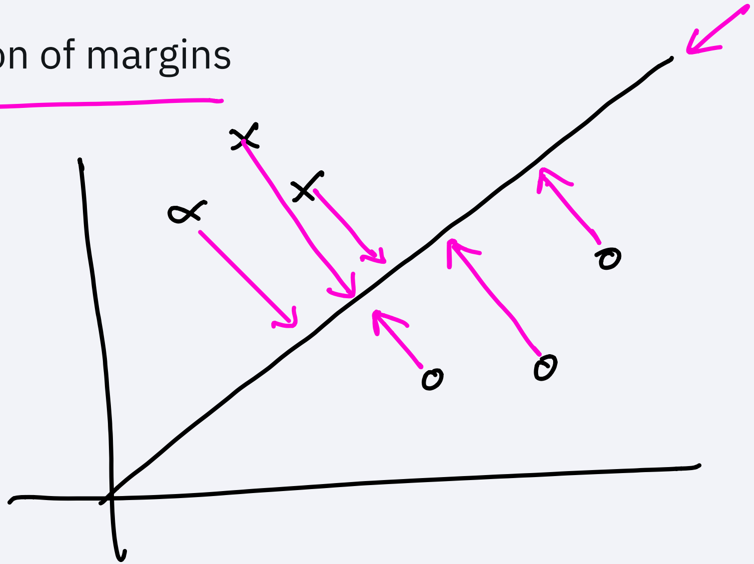


Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530.

Capacity measures

Qiskit

Distribution of margins



Capacity measures

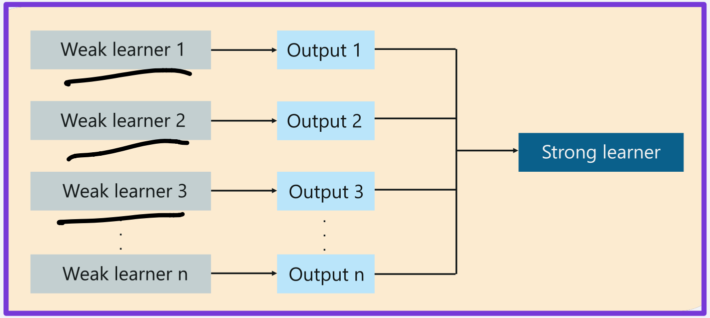
Distribution of margins

- Linear classifiers **that produce large margins** *(Vapnik, V., & Chervonenkis, A. (1974). Theory of pattern recognition.)*

Capacity measures

Distribution of margins

- Linear classifiers **that produce large margins** *(Vapnik, V., & Chervonenkis, A. (1974). Theory of pattern recognition.)*
- Explained the "paradox" observed in **boosting methods** *(Bartlett, P., Freund, Y., Lee, W. S., & Schapire, R. E. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. The annals of statistics, 26(5), 1651-1686.)*

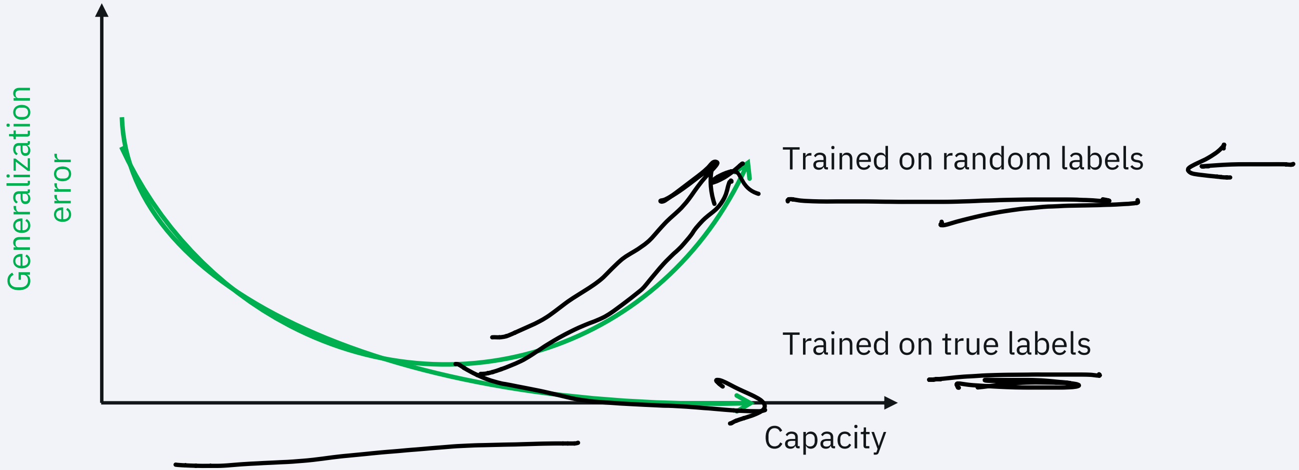


Capacity measures

- Unfortunately, margins don't seem to work with neural networks

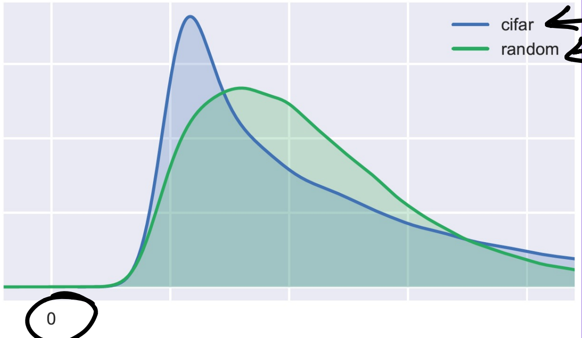
Capacity measures

- Unfortunately, margins don't seem to work with neural networks
- The margin does not inform us about generalization behavior when we consider the randomized experiment example by Zhang et al.



Capacity measures

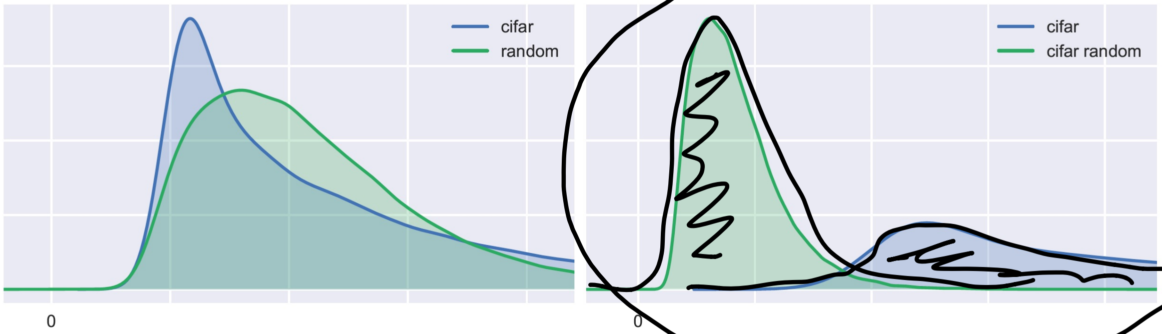
Looking at the margin alone, is not sufficient *(Bartlett, P., Foster, D. J., & Telgarsky, M. (2017). Spectrally-normalized margin bounds for neural networks. arXiv preprint arXiv:1706.08498.)*



(a) Margins.

Capacity measures

Looking at the margin alone, is not sufficient *(Bartlett, P., Foster, D. J., & Telgarsky, M. (2017). Spectrally-normalized margin bounds for neural networks. arXiv preprint arXiv:1706.08498.)*



(a) Margins. (b) Normalized margins.

Capacity measures

Qiskit

- Looking at the margin alone, is not sufficient *(Bartlett, P., Foster, D. J., & Telgarsky, M. (2017). Spectrally-normalized margin bounds for neural networks. arXiv preprint arXiv:1706.08498.)*

- PAC-Bayes approaches *(Neyshabur, B., Bhojanapalli, S., & Srebro, N. (2017). A pac-bayesian approach to spectrally-normalized margin bounds for neural networks. arXiv preprint arXiv:1707.09564.)*

- Sharpness-based measures *(Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2016). On large-batch training for deep learning: Generalization gap and sharp minima. arXiv preprint arXiv:1609.04836.)*

- Norm-based measures *(Liang, T., Poggio, T., Rakhlin, A., & Stokes, J. (2019, April). Fisher-rao metric, geometry, and complexity of neural networks. In The 22nd International Conference on Artificial Intelligence and Statistics (pp. 888-896). PMLR.)*

Capacity measures

In general, finding an appropriate capacity measure is still a relevant research question

Capacity measures

In general, finding an appropriate capacity measure is still a relevant research question

Scale-sensitive?

$y = \theta x$

$\times$ constant

$y = \sigma(\theta x)$

Capacity measures

In general, finding an appropriate capacity measure is still a relevant research question

Scale-sensitive?

Data-dependent?

incorporate
data

→ size / # params

Capacity measures

Qiskit

In general, finding an appropriate capacity measure is still a relevant research question

Scale-sensitive?

Data-dependent?

Calculable?

Capacity measures

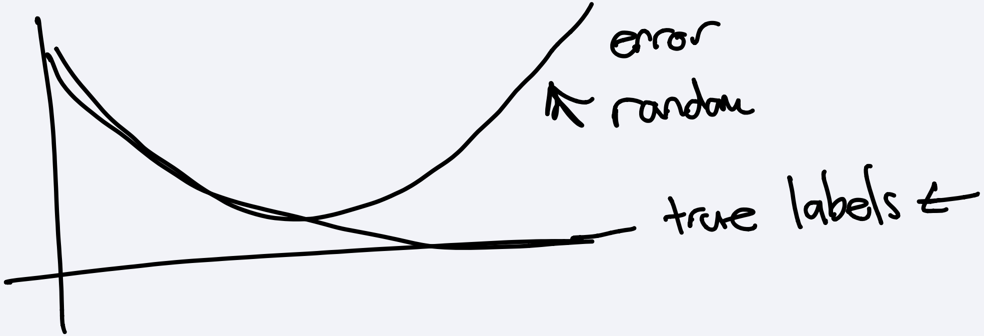
In general, finding an appropriate capacity measure is still a relevant research question

Scale-sensitive?

Data-dependent?

Calculable?

Correlates with generalization error?



Capacity of quantum models

The Capacity of Quantum Neural Networks

Logan G. Wright[1,2,*] and Peter L. McMahon[1,2,†]
1. School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA and
2. E. L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA

A key open question in quantum computation is what advantages quantum neural networks (QNNs) may have over classical neural networks (NNs), and in what situations these advantages may transpire. Here we address this question by studying the memory capacity C of QNNs, which is a metric of the expressive power of a QNN that we have adapted from classical NN theory. We present a capacity inequality showing that the capacity of a QNN is bounded by the information W that can be trained into its parameters: $C \leq W$. One consequence of this bound is that QNNs that are parameterized classically do not show an advantage in capacity over classical NNs having an equal number of parameters. However, QNNs that are parametrized with quantum states could have exponentially larger capacities. We illustrate our theoretical results with numerical experiments by simulating a particular QNN based on a Gaussian Boson Sampler. We also study the influence of sampling due to wavefunction collapse during operation of the QNN, and provide an analytical expression connecting the capacity to the number of times the quantum system is measured.

On the Quantum versus Classical Learnability of Discrete Distributions

Ryan Sweke[1], Jean-Pierre Seifert[2,3], Dominik Hangleiter[1], and Jens Eisert[1,4,5]

[1]Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, D-14195 Berlin, Germany
[2]Department of Electrical Engineering and Computer Science, TU Berlin, D-10587 Berlin, Germany
[3]FhG SIT, D-64295 Darmstadt, Germany
[4]Helmholtz Center Berlin, D-14109 Berlin, Germany
[5]Department of Mathematics and Computer Science, Freie Universität Berlin, D-14195 Berlin, Germany

Here we study the comparative power of classical and quantum learners for generative modelling within the Probably Approximately Correct (PAC) framework. More specifically we consider the following task: Given samples from some unknown discrete probability distribution, output with high probability an efficient algorithm for generating new samples from a good approximation of the original distribution. Our primary result is the explicit construction of a class of discrete probability distributions which, under the decisional Diffie-Hellman assumption, is provably not efficiently PAC learnable by a classical generative modelling algorithm, but for which we construct an efficient quantum learner. This class of distributions therefore provides a concrete example of a generative modelling problem for which quantum learners exhibit a provable advantage over classical learning algorithms. In addition, we discuss techniques for proving classical generative modelling hardness results, as well as the relationship between the PAC learnability of Boolean functions and the PAC learnability of discrete probability distributions.

The effect of data encoding on the expressive power of variational quantum machine learning models

Maria Schuld,[1] Ryan Sweke,[2] and Johannes Jakob Meyer[2]
[1]Xanadu, Toronto, ON, M5G 2C8, Canada
[2]Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany
(Dated: March 10, 2021)

Quantum computers can be used for supervised learning by treating parametrised quantum circuits as models that map data inputs to predictions. While a lot of work has been done to investigate practical implications of this approach, many important theoretical properties of these models remain unknown. Here we investigate how the strategy with which data is encoded into the model influences the expressive power of parametrised quantum circuits as function approximators. We show that one can naturally write a quantum model as a partial Fourier series in the data, where the accessible frequencies are determined by the nature of the data encoding gates in the circuit. By repeating simple data encoding gates multiple times, quantum models can access increasingly rich frequency spectra. We show that there exist quantum models which can realise all possible sets of Fourier coefficients, and therefore, if the accessible frequency spectrum is asymptotically rich enough, such models are universal function approximators.

PSEUDO-DIMENSION OF QUANTUM CIRCUITS

MATTHIAS C. CARO[†1] AND ISHAUN DATTA[‡1,2]

[1] Technical University of Munich, Germany, Department of Mathematics
nford University, USA, Institute for Computational and Mathematical Engineering

BSTRACT. We characterize the expressive power of quantum circuits with the pseudo-dimension, measure of complexity for probabilistic concept classes. We prove pseudo-dimension bounds n the output probability distributions of quantum circuits; the upper bounds are polynomial in ircuit depth and number of gates. Using these bounds, we exhibit a class of circuit output states out of which at least one has exponential gate complexity of state preparation, and moreover demonstrate that quantum circuits of known polynomial size and depth are PAC-learnable.

Generalization in Quantum Machine Learning: a Quantum Information Perspective

Leonardo Banchi,[1,2,*] Jason Pereira,[3,1] and Stefano Pirandola[3]
[1]Department of Physics and Astronomy, University of Florence, via G. Sansone 1, I-50019 Sesto Fiorentino (FI), Italy
[2]INFN Sezione di Firenze, via G. Sansone 1, I-50019, Sesto Fiorentino (FI), Italy
[3]Department of Computer Science, University of York, York YO10 5GH, UK

We study the machine learning problem of generalization when quantum operations are used to classify either classical data or quantum channels, where in both cases the task is to learn from data how to assign a certain class c to inputs x via measurements on a quantum state $\rho(x)$. A trained quantum model generalizes when it is able to predict the correct class for previously unseen data. We show that the accuracy and generalization capability of quantum classifiers depend on the (Rényi) mutual informations $I(C{:}Q)$ and $I_2(X{:}Q)$ between the quantum embedding Q and the classical input space X or class space C. Based on the above characterization, we then show how different properties of Q affect classification accuracy and generalization, such as the dimension of the Hilbert space, the amount of noise, and the amount of neglected information via, e.g., pooling layers. Moreover, we introduce a quantum version of the Information Bottleneck principle that allows us to explore the various tradeoffs between accuracy and generalization.

Connecting ansatz expressibility to gradient magnitudes and barren plateaus

Zoë Holmes,[1] Kunal Sharma,[2,3] M. Cerezo,[3,4] and Patrick J. Coles[3]
[1]Information Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA.
[2]Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA
[3]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[4]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA

Parameterized quantum circuits serve as ansätze for solving variational problems and provide a flexible paradigm for programming near-term quantum computers. Ideally, such ansätze should be highly expressive so that a close approximation of the desired solution can be accessed. On the other hand, the ansatz must also have sufficiently large gradients to allow for training. Here, we derive a fundamental relationship between these two essential properties: expressibility and trainability. This is done by extending the well established barren plateau phenomenon, which holds for ansätze that form exact 2-designs, to arbitrary ansätze. Specifically, we calculate the variance in the cost gradient in terms of the expressibility of the ansatz, as measured by its distance from being a 2-design. Our resulting bounds indicate that highly expressive ansätze exhibit flatter cost landscapes and therefore will be harder to train. Furthermore, we provide numerics illustrating the effect of expressiblity on gradient scalings, and we discuss the implications for designing strategies to avoid barren plateaus.

Rademacher complexity of noisy quantum circuits

Kaifeng Bu,[1,*] Dax Enshan Koh,[2,†] Lu Li,[3,4] Qingxian Luo,[4,5] and Yaobo Zhang[6,7]
[1]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA
[2]Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore
[3]Department of Mathematics, Zhejiang Sci-Tech University, Hangzhou, Zhejiang 310018, China
[4]School of Mathematical Sciences, Zhejiang University, Hangzhou, Zhejiang 310027, China
[5]Center for Data Science, Zhejiang University, Hangzhou, Zhejiang 310027, China
[6]Zhejiang Institute of Modern Physics, Zhejiang University, Hangzhou, Zhejiang 310027, China
[7]Department of Physics, Zhejiang University, Hangzhou, Zhejiang 310027, China

Noise in quantum systems is a major obstacle to implementing many quantum algorithms on large quantum circuits. In this work, we study the effects of noise on the Rademacher complexity of quantum circuits, which is a measure of statistical complexity that quantifies the richness of classes of functions generated by these circuits. We consider noise models that are represented by convex combinations of unitary channels and provide both upper and lower bounds for the Rademacher complexities of quantum circuits characterized by these noise models. In particular, we find a lower bound for the Rademacher complexity of noisy quantum circuits that depends on the Rademacher complexity of the corresponding noiseless quantum circuit as well as the free robustness of the circuit. Our results show that the Rademacher complexity of quantum circuits decreases with the increase in noise.

Structural risk minimization for quantum linear classifiers

Casper Gyurik 1 [*1], Dyon van Vreumingen [†1,2,3], and Vedran Dunjko [‡1]

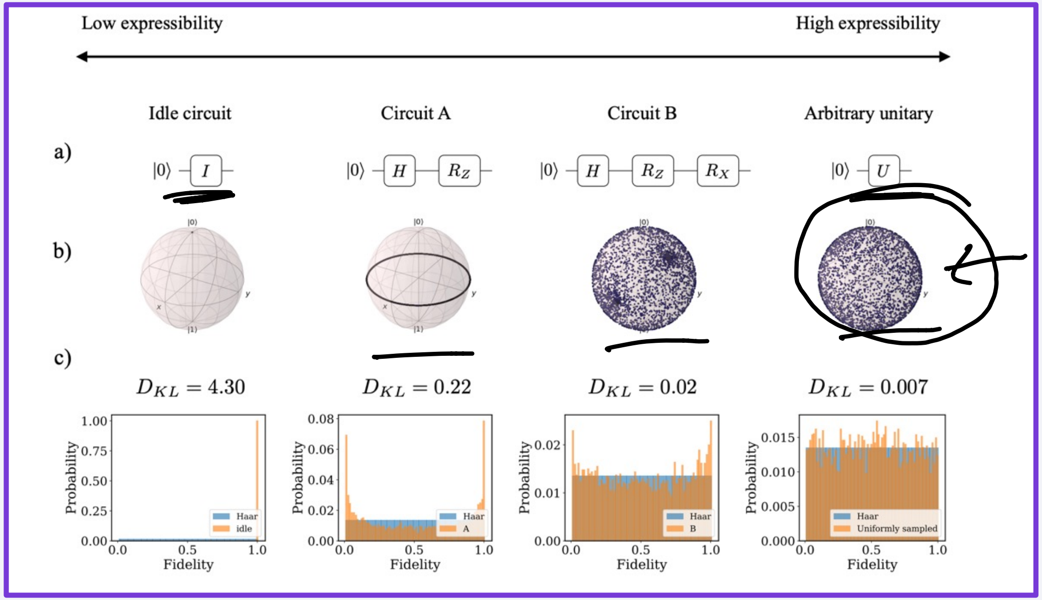
[1]Leiden University, Niels Bohrweg 1, 2333 CA Leiden, the Netherlands
[2]QuSoft, Science Park 123, 1098 XG Amsterdam, the Netherlands
titute of Physics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands

May 13, 2021

Abstract

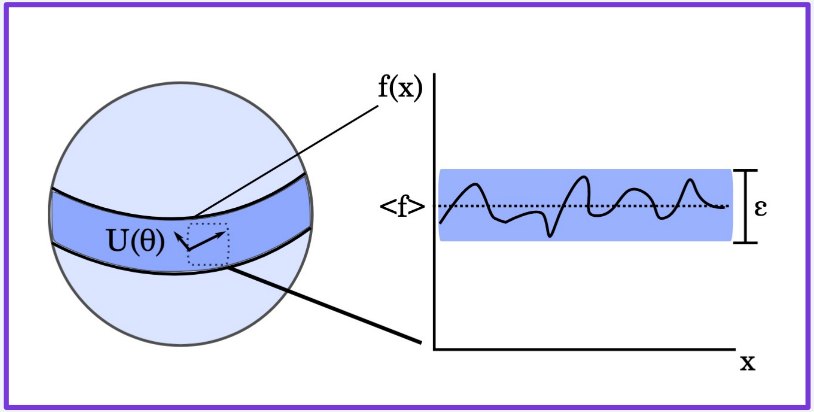
Quantum machine learning (QML) stands out as one of the typically highlighted candidates for quantum computing's near-term "killer application". In this context, QML models based on parameterized quantum circuits comprise a family of machine learning models that are well suited for implementations on near-term devices and that can potentially harness computational powers beyond what is efficiently achievable on a classical computer. However, how to best use these models – e.g., how to control their expressivity to best balance between training accuracy and generalization performance – is far from understood. In this paper we investigate capacity measures of two closely related QML models called explicit and implicit quantum linear classifiers (also called the quantum variational method and quantum kernel estimator) with the objective of identifying new ways to implement structural risk minimization – i.e., how to balance between training accuracy and generalization performance. In particular, we identify that the rank and Frobenius norm of the observables used in the QML model closely control the model's capacity. Additionally, we theoretically investigate the effect that these model parameters have on the training accuracy of the QML model. Specifically, we show that there exists datasets that require a high-rank observable for correct classification, and that there exists datasets that can only be classified with a given margin using an observable of at least a certain Frobenius norm. Our results provide new options for performing structural risk minimization for QML models.

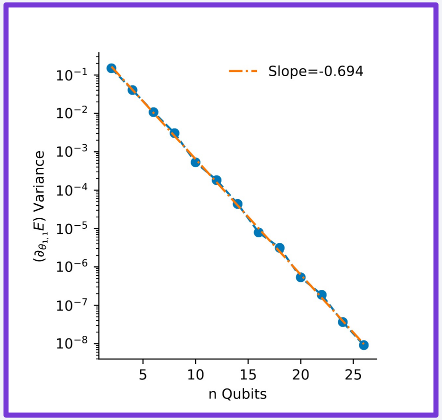
Variational circuits



Sim et al. "Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms." Advanced Quantum Technologies 2.12 (2019): 1900070.

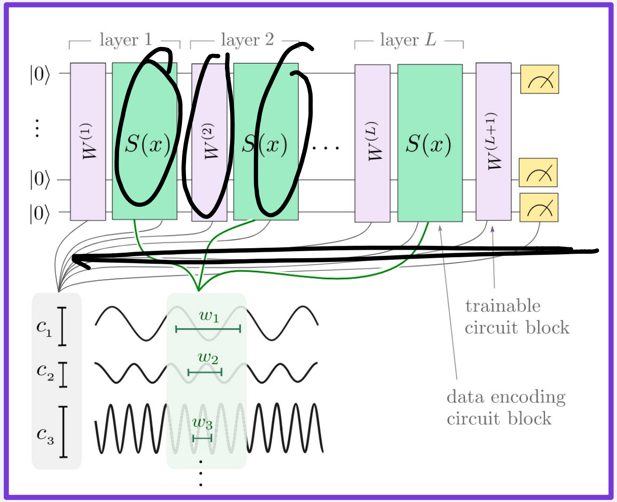
Variational circuits

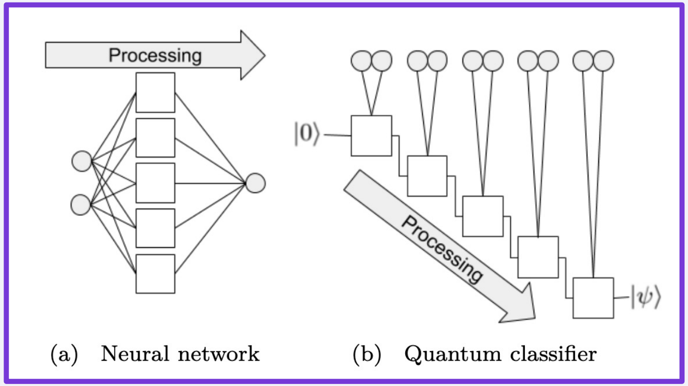




McClean, Jarrod R., et al. "Barren plateaus in quantum neural network training landscapes." Nature communications 9.1 (2018): 1-6.

Variational circuits





(a) Neural network (b) Quantum classifier

Schuld, Maria, Ryan Sweke, and Johannes Jakob Meyer. "Effect of data encoding on the expressive power of variational quantum-machine-learning models." Physical Review A 103.3 (2021): 032430.

Pérez-Salinas, Adrián, et al. "Data re-uploading for a universal quantum classifier." Quantum 4 (2020): 226.

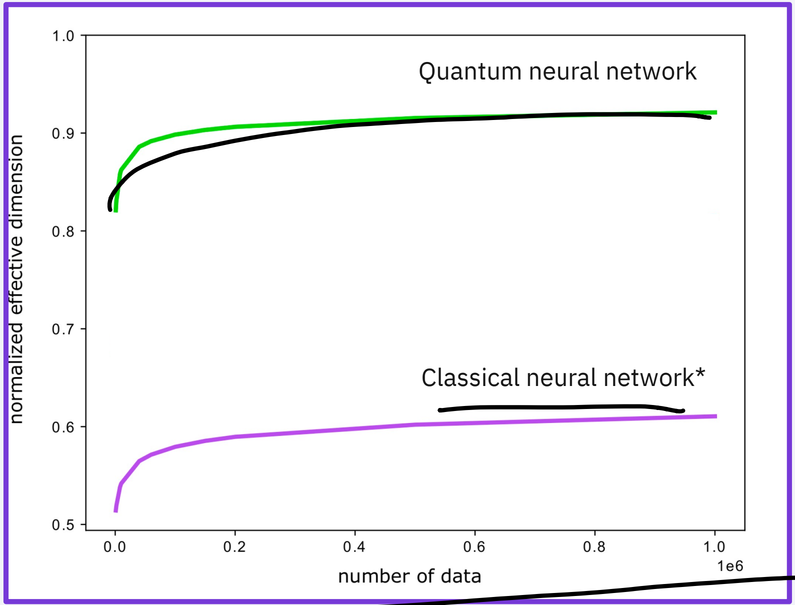
Effective dimension

1.37

Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., & Woerner, S. (2020). The power of quantum neural networks. arXiv preprint arXiv:2011.00027.

Effective dimension

Qiskit



**Chose the best possible classical ffnn out of all possible configurations*

Input size = 4

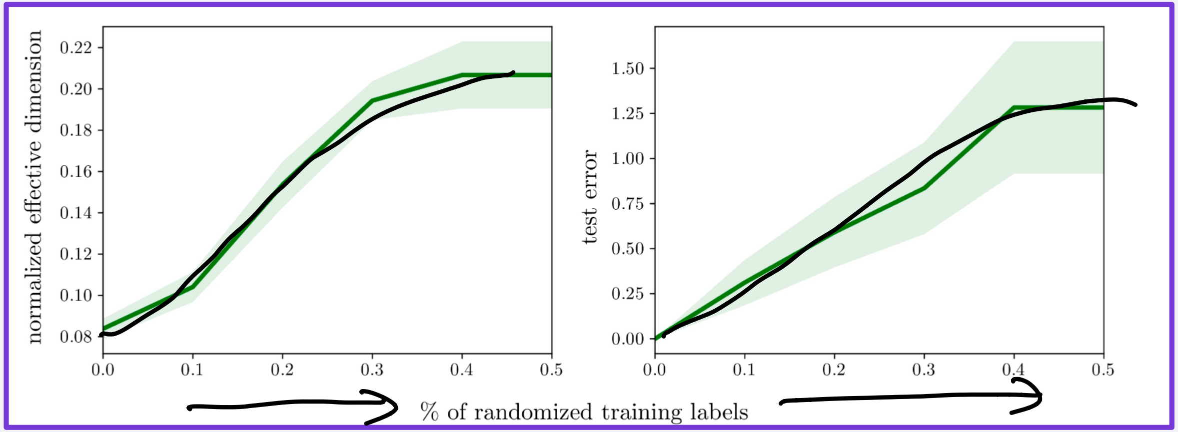
Output size = 2

$d = 40$

Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., & Woerner, S. (2020). The power of quantum neural networks. arXiv preprint arXiv:2011.00027.

Effective dimension and generalization

Found that the effective dimension for a model trained on confusion sets with increasing label corruption accurately captures generalisation behaviour



Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., & Woerner, S. (2020). The power of quantum neural networks. arXiv preprint arXiv:2011.00027.

How will QML models generalize?

Where can we hope for an advantage?

Qiskit

Capacity?	Generalization?	Data?
Computational?	Quantum kernels?	Training?
Statistical?	Variational circuits?	Applications?

Recap

Week 1:

Quantum circuits

Noise in quantum computation

Classical machine learning

Variational classifiers

QAOA

Recap

Week 1:

Quantum circuits

Noise in quantum computation

Classical machine learning

Variational classifiers

QAOA

Week 2:

Quantum feature spaces

Quantum kernels

Quantum models and training

Hardware

Advanced QML models

You made it!

Thank you!

Amira Abbas

@AmiraMorphism