Learning complexity gradually in quantum machine learning models

Erik Recio-Armengol,[1,2] Franz J. Schreiber,[3] and Carlos Bravo-Prieto[3]

[1]*ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[2]*Eurecat, Technology Center of Catalonia, Multimedia Technologies, Barcelona, Spain*
[3]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*

Introduction. In light of the spectacular success of classical machine learning and the ever-improving experimental means to perform *quantum computing*, the potential and limitations of *quantum machine learning* (QML) have become a question of widespread interest. It is known that quantum systems can perform learning tasks that are beyond the capabilities of classical computers [1–4]. At the same time, there is growing evidence that the training of quantum learning models is hindered by severe problems not present in classical machine learning. Due to their probabilistic nature, computing gradients is inherently more costly for quantum learning models. Adding to that, investigations in recent years revealed that the cost landscape of quantum learning models contains many bad minima and vast areas of vanishing gradients, so-called *Barren plateaus*, that preclude trainablity [5–10]. To overcome these issues, it is necessary to implement a favorable *inductive bias* through the model architecture or training procedure that steers the optimization problem away from these obstructions. Despite considerable research efforts in this direction, the engineering of inductive bias remains a challenging problem central to the field of quantum machine learning [11–16]. There is an urgent need for novel methods and techniques to address this challenge.

Main ideas. Traditionally, training examples are presented to the learner in random order. This holds true in both quantum and classical machine learning. However, on the classical side, there is an increasing awareness that imposing an order on the training data can speed up convergence or enable the learner to reach better local minima. In practice, the ordering is usually determined by a *scoring function*. Training starts on examples with low scores and increasingly includes data with higher scores as the training progresses. This idea comes in many different flavors and with various theoretical underpinnings, with the most well-known examples being *curriculum learning* [17] and *hard example mining* [18]. These strategies differ primarily in their approach to devising good scoring functions. In hard example mining, one aims to identify examples that are difficult to classify, expecting the learner to learn most effectively from these. The opposing viewpoint is taken in curriculum learning, where the learner progresses to hard examples only at a later stage in training, an approach inspired by human learning. Both (and other) strategies can lead to success, which is why here we take an agnostic stance on the question of design principles. Ultimately, the quality of a scoring function is decided by empirical performance. The general idea to impose an ordering upon the training data has a proven track record in a wide range of classical machine learning applications, including *natural language processing*, *computer vision* and others [18–26].

Imposing an ordering on training data has the potential to address and ease many of the challenges inherent to quantum machine learning. We show that scoring functions provide a versatile framework to incorporate inductive bias into quantum models. To this end, we draw connections to the theory of Barren plateaus and introduce a novel complexity measure for data-encoding quantum states, which gives rise to a natural scoring function to ease the problem of vanishing gradients. To our knowledge, this is the first method that addresses data-induced Barren plateaus, namely, trainability issues that arise from the input states. Generally, scoring-based training can be seen as an instance of warm-starting methods, which have taken a prominent position as candidates for addressing Barren plateaus [27–29]. We further showcase the potential of ordering training data by applying different scoring functions from curriculum learning and hard example mining to state-of-the-art applications in quantum machine learning. We observe that even straightforward strategies can lead to considerable improvements in convergence speed and overall accuracy of the quantum models.

Preliminaries. We consider a supervised learning scenario with training data $\mathbb{X} = \{X_i\}_{i=1}^N = \{|\psi_i\rangle, y_i\}_{i=1}^N$ with data-encoding quantum states $|\psi_i\rangle \in \mathcal{H}$ and labels $y \in \mathcal{Y}$ and a learning model $f_{\boldsymbol{\vartheta}} : \mathcal{H} \to \mathcal{X}$ with trainable parameters $\boldsymbol{\vartheta}$. The goal is to minimize the loss $\ell(\boldsymbol{\vartheta}; (|\psi\rangle, y))$. Traditionally, the learner has access to the whole training data $\mathbb{X}$ from the beginning of training. A common strategy is then to randomize the ordering in $\mathbb{X}$, split it into mini-batches $\mathbb{B}_i$ and perform gradient based optimization. Here, we deviate from this setting by introducing a scoring function $s : \mathcal{H} \to [0, 1]$ which assigns a score $s(|\psi\rangle)$ to every data point $|\psi_i\rangle$. From now on, we consider $\mathbb{X}$ to be ordered in ascending value of $s(|\psi_i\rangle)$ and denote with $\mathbb{X}_\alpha$ the subset that includes training examples up to index α. Additionally, we define a pacing function $p : \mathbb{N} \to \{1, 2, \ldots, N\}$. At an epoch t, the learner then has access to the training data $\mathbb{X}_{p(t)}$. Pacing functions are usually chosen to be monotonically increasing functions. Most commonly, they are chosen to be simple step function with a constant step size [19].

A common approach to obtain a scoring function is to train the learner first on the unordered dataset $\mathbb{X}$, resulting in a hypothesis $f_{\tilde{\boldsymbol{\vartheta}}}$. Then, the corresponding loss $\ell_{\tilde{\boldsymbol{\vartheta}}}$ is used as a scoring function and the training data is ordered either from low to high loss (curriculum learning) or from high to low (hard example mining). A similar, less costly method is so-called *self-paced learning*. During training, the hypothesis $f_{\boldsymbol{\vartheta}}$ is continually updated. In self-paced learning, the loss $\ell_{\boldsymbol{\vartheta}}$ corresponding to the current hypothesis $f_{\boldsymbol{\vartheta}}$ is used to order the training data for the next epoch. Another approach is to first train another, simpler classifier $g_{\boldsymbol{\vartheta}}$, for example a *support vector machine*, on $\mathbb{X}$ and use again the corresponding loss function $\ell_{\boldsymbol{\vartheta}}$ as a scoring function. We also explore an innovative approach leveraging inherently quantum scoring functions. Specifically, we investigate the use of the $\mathfrak{g}$-purity $\mathcal{P}_{\mathfrak{g}}(|\psi_i\rangle)$ of training states, a metric known to influence the occurrence of barren plateaus in quantum machine learning [5]. By ordering the training set based on this purity, we aim to mitigate the detrimental effects of barren plateaus and improve model performance.

Results. Due to space limitations, we highlight the application of our framework to quantum phase recognition with self-paced learning. We demonstrate its effectiveness on two paradigmatic quantum spin chains: the generalized cluster model [30] and the bond-alternating XXZ model [31]. Their ground-state phase diagrams are presented in Fig. 1.

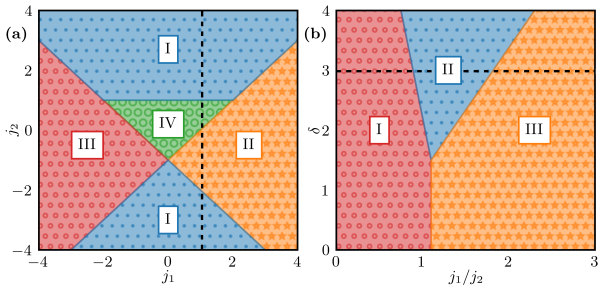

Figure 1. Ground-state phase diagrams of **(a)** the generalized cluster Hamiltonian, exhibiting (I) symmetry-protected topological, (II) ferromagnetic, (III) antiferromagnetic, and (IV) trivial phases; and **(b)** the bond-alternating XXZ Hamiltonian, displaying (I) trivial, (II) symmetry-broken antiferromagnetic, and (III) topological phases.

We leverage the quantum convolutional neural network (QCNN) architecture introduced in Ref. [32] for the phase classification tasks. Our QCNN maps an n-qubit input state vector $|\psi\rangle$ into a 2-qubit output state, where the predicted label $\hat{\boldsymbol{y}}$ is derived from the probabilities measured in the computational basis:

$$|\psi\rangle \mapsto (p_j)_{j\in\{0,1\}^2} \mapsto \hat{\boldsymbol{y}} := (p_{00}, p_{01}, p_{10}, p_{11})\,.$$

Each component of $\hat{\boldsymbol{y}}$ represents a distinct phase class. In our experiments, we utilize 50 ground states for training and 1000 samples from the same distribution for testing.

Self-paced learning. Self-paced learning dynamically prioritizes training data based on the current model's performance. At each epoch, we rank the training set by the loss of each data point. With minibatches of size 10, we explore four strategies using a monotonically increasing pacing function, which gradually exposes more data points as training progresses: (1) *Vanilla*; standard training with no specific ordering nor pacing function, (2) *Easy*; prioritizes points with lower loss values (easier examples), (3) *Hard*; prioritizes points with higher loss values (harder examples), and (4) *Hardest*; employs a constant pacing function and always trains on the 10 most difficult examples. As shown in Fig. 2, the Vanilla and Easy strategies yield comparable performance, reaching a final accuracy of nearly 80%. The Hard strategy initially performs well but regresses to a similar level as more data points are

included (due to the monotonically increasing pacing function). In stark contrast, the Hardest strategy, focusing solely on the most difficult examples, rapidly converges to near-perfect accuracy on the training set and over 90% on the test set. This highlights the inductive bias introduced by a specific training data order, demonstrating that the hardest-to-classify examples are the most informative for quantum phase recognition. This finding may not be the case for other strategies and machine learning tasks, as easier examples may sometimes be more informative.

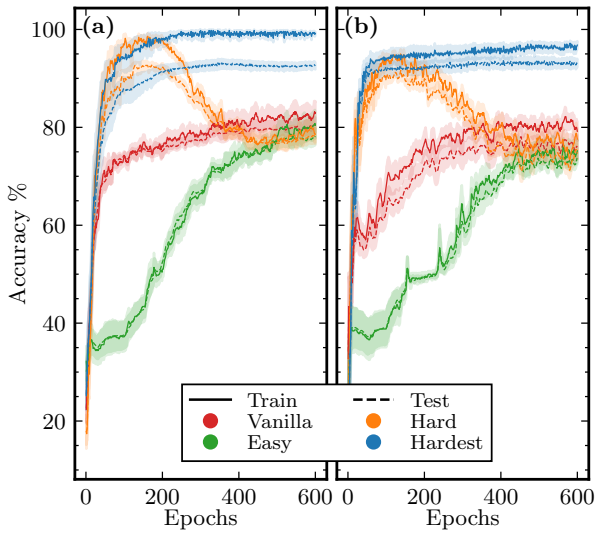

Figure 2. (a) Generalized cluster Hamiltonian and (b) alternating-bond XXZ Hamiltonian

Accuracy at the cost of confidence. To further understand the differences between the Vanilla and Hardest self-paced learning strategies, we zoom-in cut along the j_2 coupling constant (fixing $j_1 = 1$) and j_1/j_2 (fixing $\delta = 3$) for the respective Hamiltonians (dashed lines in Fig. 1). Fig. 3 shows the probabilities of measuring each basis vector along these lines, with the highest probability determining the predicted phase. The Hardest approach demonstrates superior accuracy, both within each phase region and at the boundaries between phases. Interestingly, however, the Hardest approach exhibits less confidence in its predictions, as indicated by the smaller differences between the probabilities of different classes. This suggests a potential trade-off between accuracy and confidence in this specific context.

Summary and outlook. The trainability of quantum machine learning models is a pressing challenge due to the prevalence of barren plateaus and the high cost of gradient computation. This submission introduces an approach to address these issues by strategically ordering the training data. This method, which gradually adjusts the complexity of the training data, can be seen as incorporating an inductive bias into quantum models, leading to improved convergence speed and overall accuracy. While this extended abstract focuses on a specific strategy and quantum phase recognition task, our approach broadly applies to other strategies and QML tasks, such as unitary learning or quantum error correction. We consider this submission to underscore the potential of introducing inductive bias through the training data as a powerful tool for mitigating the challenges of barren plateaus and enhancing trainability of QML models.

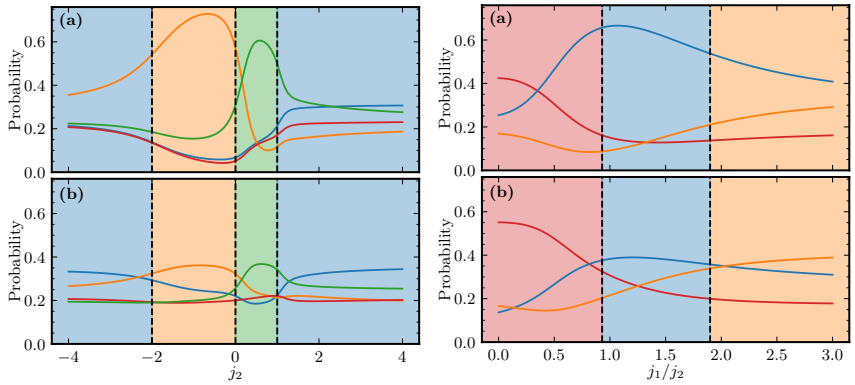

Figure 3. (Left) Generalized Cluster Hamiltonian and (Right) bond-alternating XXZ Hamiltonian for (a) Vanilla and (b) Hardest. The background color of each panel indicates the true quantum phase of the system. A correct classification occurs when the highest probability color matches the background color of the panel.

[1] N. Pirnay, R. Sweke, J. Eisert, and J.-P. Seifert, Superpolynomial quantum-classical separation for density modeling, Physical Review A **107**, 042416 (2023).

[2] R. Sweke, J.-P. Seifert, D. Hangleiter, and J. Eisert, On the quantum versus classical learnability of discrete distributions, Quantum **5**, 417 (2021).

[3] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, Nature Physics **17**, 1013 (2021).

[4] H.-Y. Huang, R. Kueng, and J. Preskill, Information-theoretic bounds on quantum advantage in machine learning, Physical Review Letters **126**, 190505 (2021).

[5] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. O. Marrero, M. Larocca, and M. Cerezo, A unified theory of barren plateaus for deep parametrized quantum circuits, arXiv preprint arXiv:2309.09342 (2023).

[6] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature communications **9**, 4812 (2018).

[7] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, A review of barren plateaus in variational quantum computing, arXiv preprint arXiv:2405.00781 (2024).

[8] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, Nature Communications **13**, 7760 (2022).

[9] N. A. Nemkov, E. O. Kiktenko, and A. K. Fedorov, Barren plateaus are swamped with traps, arXiv preprint arXiv:2405.05332 (2024).

[10] S. Thanasilp, S. Wang, N. A. Nghiem, P. Coles, and M. Cerezo, Subtleties in the trainability of quantum machine learning models, Quantum Machine Intelligence **5**, 21 (2023).

[11] J. Bowles, V. J. Wright, M. Farkas, N. Killoran, and M. Schuld, Contextuality and inductive bias in quantum machine learning, arXiv preprint arXiv:2302.01365 (2023).

[12] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, Exploiting symmetry in variational quantum machine learning, PRX Quantum **4**, 010328 (2023).

[13] S. Jerbi, L. J. Fiderer, H. Poulsen Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, Quantum machine learning beyond kernel methods, Nature Communications **14**, 517 (2023).

[14] J. Kübler, S. Buchholz, and B. Schölkopf, The inductive bias of quantum kernels, Advances in Neural Information Processing Systems **34**, 12661 (2021).

[15] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, Nature Computational Science **2**, 567 (2022).

[16] M. Larocca, F. Sauvage, F. M. Sbahi, G. Verdon, P. J. Coles, and M. Cerezo, Group-invariant quantum machine learning, PRX Quantum **3**, 030341 (2022).

[17] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, Curriculum learning, in *Proceedings of the 26th annual international conference on machine learning* (2009) pp. 41–48.

[18] A. Shrivastava, A. Gupta, and R. Girshick, Training region-based object detectors with online hard example mining, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 761–769.

[19] X. Wang, Y. Chen, and W. Zhu, A survey on curriculum learning, IEEE Transactions on Pattern Analysis and Machine Intelligence **44**, 4555 (2021).

[20] P. Soviany, R. T. Ionescu, P. Rota, and N. Sebe, Curriculum learning: A survey, International Journal of Computer Vision **130**, 1526 (2022).

[21] R. Tudor Ionescu, B. Alexe, M. Leordeanu, M. Popescu, D. P. Papadopoulos, and V. Ferrari, How hard can it be? estimating the difficulty of visual search in an image, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) pp. 2157–2166.

[22] S. Guo, W. Huang, H. Zhang, C. Zhuang, D. Dong, M. R. Scott, and D. Huang, Curriculumnet: Weakly supervised learning from large-scale web images, in *Proceedings of the European conference on computer vision (ECCV)* (2018) pp. 135–150.

[23] L. Jiang, D. Meng, T. Mitamura, and A. G. Hauptmann, Easy samples first: Self-paced reranking for zero-example multimedia search, in *Proceedings of the 22nd ACM international conference on Multimedia* (2014) pp. 547–556.

[24] E. A. Platanios, O. Stretcu, G. Neubig, B. Poczos, and T. M. Mitchell, Competence-based curriculum learning for neural machine translation, arXiv preprint arXiv:1903.09848 (2019).

[25] T. Matiisen, A. Oliver, T. Cohen, and J. Schulman, Teacher–student curriculum learning, IEEE transactions on neural networks and learning systems **31**, 3732 (2019).

[26] X. Zhang, G. Kumar, H. Khayrallah, K. Murray, J. Gwinnup, M. J. Martindale, P. McNamee, K. Duh, and M. Carpuat, An empirical exploration of curriculum learning for neural machine translation, arXiv preprint arXiv:1811.00739 (2018).

[27] A. A. Mele, G. B. Mbeng, G. E. Santoro, M. Collura, and P. Torta, Avoiding barren plateaus via transferability of smooth solutions in a hamiltonian variational ansatz, Physical Review A **106**, L060401 (2022).

[28] M. Drudis, S. Thanasilp, Z. Holmes, *et al.*, Variational quantum simulation: a case study for understanding warm starts, arXiv preprint arXiv:2404.10044 (2024).

[29] M. Cerezo, M. Larocca, D. García-Martín, N. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, *et al.*, Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing, arXiv preprint arXiv:2312.09121 (2023).

[30] R. Verresen, R. Moessner, and F. Pollmann, One-dimensional symmetry protected topological phases and their transitions, Phys. Rev. B **96**, 165124 (2017).

[31] A. Elben, J. Yu, G. Zhu, M. Hafezi, F. Pollmann, P. Zoller, and B. Vermersch, Many-body topological invariants from randomized measurements in synthetic quantum matter, Science advances **6**, eaaz3666 (2020).

[32] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nature Physics **15**, 1273 (2019).