

Local tensor-train surrogates for quantum learning models

Sreeraj Rajindran Nair^{1 *}

Christopher Ferrie^{1 †}

¹ *Centre for Quantum Software and Information, University of Technology Sydney, NSW 2007, Australia*

Abstract. A key bottleneck in quantum machine learning is the computational cost of repeated quantum circuit evaluations during the inference phase. To address this, we present a framework for constructing fast, cheap, provably accurate classical tensor-train surrogates of fully trained quantum machine learning models within local patches of their input data space. The approach combines Taylor polynomial approximation with a tensor-train (TT) representation and embeds it in a statistical learning paradigm via empirical risk minimization. In our analysis, the Taylor-TT construction serves as a deterministic error certificate proving that the TT hypothesis class contains a good approximation; empirical risk minimization then provably recovers a surrogate with controlled generalization error and explicit bounds. This translates into three independently controllable error sources: (i) Taylor truncation error controlled by the patch radius r and polynomial degree p , (ii) TT approximation error controlled by the bond dimension χ , and (iii) statistical estimation error. While the parameter count scales polynomially in the number of data dimensions N , i.e., $= N(p+1)\chi^2$ rather than the naive $(p+1)^N$, the worst-case constants inherit an exponential factor through the tensor-product feature norm during Taylor polynomial embedding onto TT. This cleanly separates representation complexity from feature-induced constants. Our risk bounds and sample complexity depend explicitly on the local patch radius r . The arXiv preprint with the complete theory with proofs and numerical experiments is available at <https://arxiv.org/abs/2604.25631>.

Keywords: Quantum machine learning, tensor-train decomposition, PAC learning

1 Significance for quantum technology

Quantum machine learning (QML) models, particularly parametrized quantum circuits, have shown promise for learning tasks on near-term and fault-tolerant quantum hardware. However, a fundamental practical challenge lies not in training these models, but in *deploying* them: every prediction, gradient evaluation, sensitivity query, or interpretability analysis requires executing a quantum circuit. Unlike classical models, which, once trained, can be queried at negligible cost on classical hardware, quantum models impose a per-query cost in quantum hardware time, energy, and expense that scales with the circuit’s complexity.

Classical surrogates address this: after training on quantum hardware (where potential quantum advantages may be harvested) the trained quantum model is replaced by a classical function that reproduces its behaviour. Previous work has explored *global* surrogation, constructing replacements valid across the entire input domain [1, 2]. We pursue a complementary strategy: *local* surrogation, building a classical surrogate valid within a targeted local patch [3] of the input space, extending our prior numerical work [4]. Aside from dequantizing the inference phase, local surrogates are also relevant for post-training tasks such as interpretability [5], sensitivity analysis, and local optimization, where one cares about model behavior in a specific neighborhood rather than globally.

2 The framework

We consider a trained quantum model as a black box function $g(x)$ and a local patch $B(x_0, r)$ around a point of interest. We define a *learning task* as a regression problem, where the learner collects input-output samples (X_i, Y_i) from the patch (where $Y_i = g(X_i) + \varepsilon_i$ may include measurement noise) and fits a predictor from a constrained tensor-train (TT) hypothesis class that minimizes the expected squared loss via empirical risk minimization (ERM) [6, 7]. The hypothesis class $\mathcal{H}_{\text{TT}}(\Lambda, \chi)$ consists of functions of the form $h_A(x) = \langle A, \Phi(\xi) \rangle$, where Φ is a degree- p tensor-product feature map, $\xi = (x - x_0)/r$ are normalized coordinates, and A is a tensor constrained to have tensor-train (TT) rank at most χ and Frobenius norm

*sreeraj.r.nair@student.uts.edu.au

†christopher.ferrie@uts.edu.au

at most Λ . The theoretical backbone of our framework is a *deterministic error certificate* constructed via Taylor approximation and TT representation. For this, we construct the degree- p Taylor polynomial of g at x_0 . After expressing this polynomial using a factorial feature map and embedding the total-degree coefficients into a $(p+1)^N$ box tensor $A_\star$, we apply the TT-SVD decomposition [8] to obtain a rank- χ approximation A_χ . The resulting predictor h_{A_χ} satisfies, for all $x \in B(x_0, r)$,

$$|g(x) - h_{A_\chi}(x)| \leq \underbrace{\varepsilon_{\text{Taylor}}(r, p)}_{\text{Taylor truncation}} + \underbrace{K^N \|A_\star - A_\chi\|_F}_{\text{TT approximation}} =: E_{\text{det}}.$$

Here $K \approx 1.51$ is a Bessel constant arising from the feature norm bound and the TT approximation error depends on the singular value decay of the tensor’s unfoldings and can be driven to zero by increasing χ . The predictor h_{A_χ} lives inside the hypothesis class $\mathcal{H}_{\text{TT}}(\Lambda, \chi)$. This acts as an *existence proof for the richness of the hypothesis class* providing guarantees that the TT hypothesis class contains a good approximation to g within E_{det} , and that ERM can converge onto it. In practice, the learner does not need to compute Taylor coefficients or perform TT-SVD. With the existence of a good hypothesis established, the statistical analysis proceeds by using standard theoretical machine learning tools to bound the population/true risk. We use uniform convergence [9] of empirical risk to population risk to show that the ERM solution generalizes i.e. low training error implies low generalization error. The key steps in our analytical proof are:

1. We bound the *pseudo-dimension* of the $\mathcal{H}_{\text{TT}}(\Lambda, \chi)$ to quantify the hypothesis class complexity. Leveraging tensor-network parameter counting results [10], the pseudo-dimension scales as $d_{\text{eff}} = O(N(p+1)\chi^2 \log N)$ i.e. polynomially in the input dimension N , the polynomial degree p , and the bond dimension χ , in contrast to the exponential $(p+1)^N$ scaling of an unconstrained polynomial model.
2. Pseudo-dimension control yields *uniform convergence*: with high probability over the training sample, the empirical risk of every hypothesis in the class is close to its population risk simultaneously. This is a standard result in statistical learning theory, applied here to the TT-constrained class [11, 12].
3. Combining these, the ERM solution $\hat{h}$ satisfies, with probability $\geq 1 - \delta$, the population risk $R(\hat{h})$ (expected squared error on unseen points) is bounded as:

$$R(\hat{h}) \leq E_{\text{det}}^2 + \Delta_n(\delta),$$

where $\Delta_n(\delta) \rightarrow 0$ as $n \rightarrow \infty$. The certificate E_{det}^2 enters as an upper bound on the *best achievable risk* in the class; ERM finds a hypothesis at most Δ_n worse.

An important subtlety is that while the *parameter count* d_{eff} scales polynomially, the *worst-case constants* in the bounds inherit an exponential factor K^N through the tensor-product feature norm. As a consequence, the sample complexity scales as K^{4N} , which could be improved upon in future work by utilizing more efficient TT constructions. This separates *representation complexity* (polynomial via TT structure) from *feature-induced constants* (exponential via the embedding), clarifying precisely where the curse of dimensionality enters.

In summary, we propose a model agnostic analytical framework for local TT surrogates wherein the total approximation error decomposes into three transparent sources, each independently controllable:

- *Polynomial truncation error*, controlled by the patch radius r and polynomial degree p ;
- *Tensor compression error*, controlled by the TT bond dimension χ ;
- *Statistical estimation error*, arising when the surrogate is learned from finite noisy samples.

Neither the deterministic approximation component nor the statistical component alone suffices: the Taylor-TT certificate without ERM gives no learning algorithm, and ERM without the certificate gives no guarantee that the hypothesis class is rich enough. The final learned surrogate is a TT-format function but is *not* the certificate predictor. ERM is free to find any predictor in the hypothesis class, and may outperform the certificate. The resulting surrogate can be evaluated, differentiated, and integrated at negligible classical cost.

Numerical experiments on quantum convolutional neural networks trained on real-world datasets validate the theoretical predictions while using the Taylor-TT certificate as a warm start. An arXiv preprint presenting the complete framework, all proofs, and numerical experiments is available at <https://arxiv.org/abs/2604.25631>.

References

- [1] F. J. Schreiber, J. Eisert, and J. J. Meyer, “Classical surrogates for quantum learning models,” *Physical Review Letters*, vol. 131, Sept. 2023.
- [2] S. Jerbi, C. Gyurik, S. C. Marshall, R. Molteni, and V. Dunjko, “Shadows of quantum machine learning,” *Nature Communications*, vol. 15, July 2024.
- [3] S. Lerch, R. Puig, M. S. Rudolph, A. Angrisani, T. Jones, M. Cerezo, S. Thanasilp, and Z. Holmes, “Efficient quantum-enhanced classical simulation for patches of quantum landscapes,” 2024.
- [4] S. R. Nair and C. Ferrie, “Local surrogates for quantum machine learning,” 2025.
- [5] C. Molnar, *Interpretable Machine Learning*. 3 ed., 2025.
- [6] V. Vapnik, “An overview of statistical learning theory,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.
- [7] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- [8] I. V. Oseledets, “Tensor-train decomposition,” *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [9] S. Shalev-Shwartz, O. Shamir, N. Srebro, and K. Sridharan, “Learnability, stability and uniform convergence,” *Journal of Machine Learning Research*, vol. 11, no. 90, pp. 2635–2670, 2010.
- [10] B. Khavari and G. Rabusseau, “Lower and upper bounds on the pseudo-dimension of tensor network models,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- [11] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. MIT Press, 2 ed., 2018.
- [12] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer: New York, 1999.