Quantum adiabatic machine learning with zooming

Alexander Zlokapa[1], Alex Mott[2], Joshua Job[3], Jean-Roch Vlimant[1], Daniel Lidar[4], and Maria Spiropulu[1]

[1] *Division of Physics, Mathematics & Astronomy, Alliance for Quantum Technologies, California Institute of Technology, Pasadena, CA 91125, United States*
[2] *DeepMind Technologies, London, UK*
[3] *Lockheed Martin Corporation, El Segundo, CA 90245, USA*
[4] *Departments of Electrical and Computer Engineering, Chemistry, and Physics & Astronomy, and Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, CA 90089, USA*

Abstract

Recent work has shown that quantum annealing for machine learning (QAML) can perform comparably to state-of-the-art machine learning methods with a specific application to Higgs boson classification. We propose a variant algorithm (QAML-Z) that iteratively zooms in on a region of the energy surface by mapping the problem to a continuous space and sequentially applying quantum annealing to an augmented set of weak classifiers. Results on a programmable quantum annealer show that QAML-Z increases the performance difference between QAML and classical deep neural networks by over 40% as measured by area under the ROC curve for small training set sizes. Furthermore, QAML-Z reduces the advantage of deep neural networks over QAML for large training sets by around 50%, indicating that QAML-Z produces stronger classifiers that retain the robustness of the original QAML algorithm.

1 Background

In the original quantum annealing for machine learning (QAML) algorithm[1], a training set with S examples of labeled data $\{\mathbf{x}_\tau, y_\tau\}$ (where $\mathbf{x}_\tau$ is an input vector and $y_\tau = \pm 1$ is a binary label for signal and background) is optimized with a set of N weak classifiers c_i, each of which gives $c_i(\mathbf{x}_\tau) = \pm 1/N$ for a signal or background prediction. For convenience, we define the variables:

$$C_{ij} = \sum_{\tau=1}^{S} c_i(\mathbf{x}_\tau) c_j(\mathbf{x}_\tau), \qquad C_i = \sum_{\tau=1}^{S} c_i(\mathbf{x}_\tau) y_\tau. \tag{1}$$

In the original QAML algorithm, the following Ising model Hamiltonian is minimized[2]:

$$H = \sum_{i=1}^{N} \left(\lambda - C_i + \frac{1}{2} \sum_{j>i}^{N} C_{ij} \right) s_i + \frac{1}{4} \sum_{i=1}^{N} \sum_{j>i}^{N} C_{ij} s_i s_j, \tag{2}$$

where the spin of each qubit is $s_i = \pm 1$ and a regularization constant $\lambda > 0$ is chosen. The strong classifier constructed by the weak classifiers is a sum of the classifiers weighted by spin, normalizing $\{-1, 1\}$ to $\{0, 1\}$:

$$R(\mathbf{x}_\tau) = \sum_{i=1}^{N} \frac{1}{2}(s_i + 1) c_i(\mathbf{x}_\tau). \tag{3}$$

This construction of a strong classifier only enables the weak classifiers to be turned on or off, creating a limitation in the strength of the final classifier.

2 QAML-Z Algorithm

By iteratively performing quantum annealing, the binary weights on the weak classifiers can be made continuous, ultimately resulting in a stronger classifier. This is achieved by performing a search on the real numbers, effectively zooming in on a region of the energy surface each iteration. Hence, we denote the zooming variant of quantum annealing for machine learning as QAML-Z. Under this reformulation, the weights of the classifiers may be extended from the set $\{0, 1\}$ to the continuous interval $[-1, 1]$, enabling the subtraction of classifiers to reduce cross-correlations between weak classifiers.

Let each qubit have a mean $\mu_i(t)$ (starting at $\mu_i(0) = 0$ for all i) and let the search breadth be $\sigma(t) = b^t$, where $t = 0, 1, ..., T-1$ for T iterations and $0 < b < 1$ is a free parameter. We then modify the Hamiltonian by substituting $s_i c_i(\mathbf{x}_\tau) \to \sigma(t) s_i c_i(\mathbf{x}_\tau) + \mu_i(t) c_i(\mathbf{x}_\tau)$ to zoom into the region of interest, shifting and narrowing the optimization problem. To derive the Ising model energy, the distance between predictions and data is minimized:

$$||\mathbf{y} - \mathbf{R}||^2 = \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i c_i(\mathbf{x}_\tau) \right|^2. \tag{4}$$

Applying the substitution, dropping terms independent of spin, and simplifying with the C_i and C_{ij} variables previously defined, we find the following Hamiltonian:

$$H(t) = \sum_{i=1}^{N}\left(-C_i + \sum_{j>i}^{N}\mu_j(t)C_{ij}\right)\sigma(t)s_i + \sum_{i=1}^{N}\sum_{j>i}^{N}C_{ij}\sigma^2(t)s_is_j. \tag{5}$$

This new Hamiltonian may be iteratively optimized for $t = 0, 1, \ldots, T-1$ to update $\mu_i(t+1) = \mu_i(t) + s_i\sigma(t+1)$, resulting in the new strong classifier:

$$R(\mathbf{x}_\tau) = \sum_{i=1}^{N}\mu_i(T-1)c_i(\mathbf{x}_\tau). \tag{6}$$

Since the zooming algorithm increases the possibility of overfitting, we propose a two-step randomization procedure to regularize the iterative process. After each iteration, for each qubit s_i such that the energy worsens by the update (i.e., $E[\mu_0(t+1), \ldots, \mu_i(t+1), \ldots, \mu_N(t+1)] > E[\mu_0(t+1), \ldots, \mu_i(t), \ldots \mu_N(t+1)]$), we apply the flip $s_i \to -s_i$ with monotonically decreasing probability $p_f(t)$. Subsequently, all qubits are uniformly randomly flipped from s_i to $-s_i$ with probability $q_f(t)$ where $q_f(t) < p_f(t)$ for all t. This both prevents the strong classifier from overfitting as well as pushing it out of local minima in an annealing-inspired procedure. The functions p_f and q_f are specified in the SM.

To take full advantage of these continuous weights, we augment the set of original weak classifiers $h_i(\mathbf{x}_\tau)$ that returns a value in $[-1, 1]$. For each h_i, multiple classifiers are generated by shifting the threshold to round to ± 1:

$$c_{il}(\mathbf{x}_\tau) = \mathrm{sgn}(h_i(\mathbf{x}_\tau) + \delta l)/N, \tag{7}$$

where N is the number of classifiers, $l \in \mathbb{Z} : -A \leq l \leq A$ is the offset and δ is the step size. Hence, the Hamiltonian is now given by:

$$H(t) = \sum_{l=-A}^{A}\left[\sum_{i=1}^{N}\left(-C_{il} + \sum_{j>i}^{N}\mu_{jl}(t)C_{ijl}\right)\sigma(t)s_{il} + \sum_{i=1}^{N}\sum_{j>i}^{N}C_{ijl}\sigma^2(t)s_{il}s_{jl}\right], \tag{8}$$

where $H(t)$ is iteratively optimized for $t = 0, 1, \ldots$ to update $\mu_{il}(t+1) = \mu_{il}(t) + s_{il}\sigma(t+1)$. Similarly to before, we have defined:

$$C_{ijl} = \sum_{\tau=1}^{S}c_{il}(\mathbf{x}_\tau)c_{jl}(\mathbf{x}_\tau), \qquad C_{il} = \sum_{\tau=1}^{S}c_{il}(\mathbf{x}_\tau)y_\tau. \tag{9}$$

3 Application of QAML-Z to the Higgs optimization problem

As an application of QAML-Z, we revisited the Higgs optimization problem.[1] We implemented the QAML-Z algorithm on the programmable quantum annealer at the University of Southern California built by D-Wave Systems, Inc.[3] Because the D-Wave 2X architecture is not fully connected, we modified the Ising model to encourage sparsity for minor embedding on the D-Wave machine. In the augmentation scheme defined above, we set an offset of $A = 3$ and a step size of $\delta = 0.0075$. Additionally, we set the zoom parameter $b = \frac{1}{2}$ to perform a binary search over the real numbers. The weights of cross-terms in the J matrix were pruned, keeping only the largest 5% of weights nonzero. Additionally, (polynomial-time) variable-fixing procedures in the D-Wave API were used to reduce the size of the Ising model encoded on the annealer. As in the original paper,[1] each annealing consisted of averaging multiple gauges[4] and excited states, where the number of gauges, number of excited states and chain strength[5] are decayed monotonically with each iteration. Due to the definition of $\sigma(t)$, the marginal impact of each iteration follows an exponential decay, and thus QAML-Z was trained for only 8 iterations.

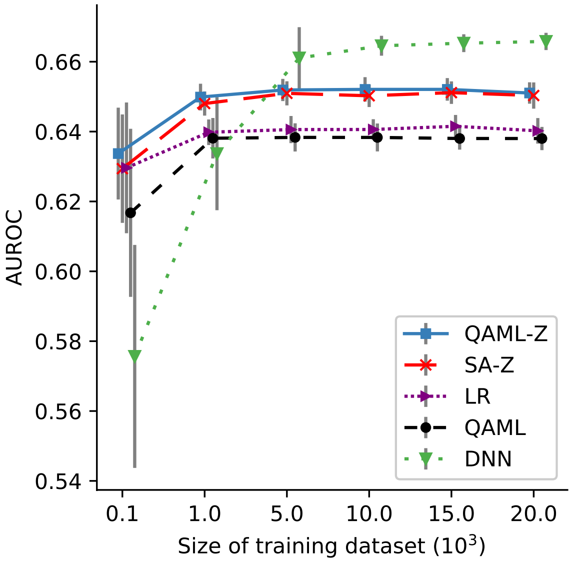

Figure 1: Area under the ROC curve for the QAML-Z extension, simulated annealing (SA-Z), the original QAML, a logistic regression (LR) and a deep neural network (DNN) as a function of training set size. The proposed QAML-Z algorithms yields an increase of 41% in the advantage of the original QAML algorithm[1] over the DNN at the smallest training set size. At the largest training set size, the disadvantage of QAML relative to the DNN is reduced by 47% by QAML-Z. Error bars indicate 1σ error, including both variation over training sets and statistical error.

Compared to the QAML algorithm, the area under the receiver operating characteristic curve (AUROC) is significantly improved by QAML-Z on all training set sizes (Figure 1). We select the best-performing traditional classifier (a deep neural network) from the original publication as a benchmark. A logistic regression directly optimizes the mean-squared error given in Eq. (4) and is thus also shown. When compared to classical simulated annealing, QAML-Z performs significantly better for all training set sizes due to its use of excited states in the neighborhood of the ground state, marking further improvement over the original QAML algorithm (Figure 2).

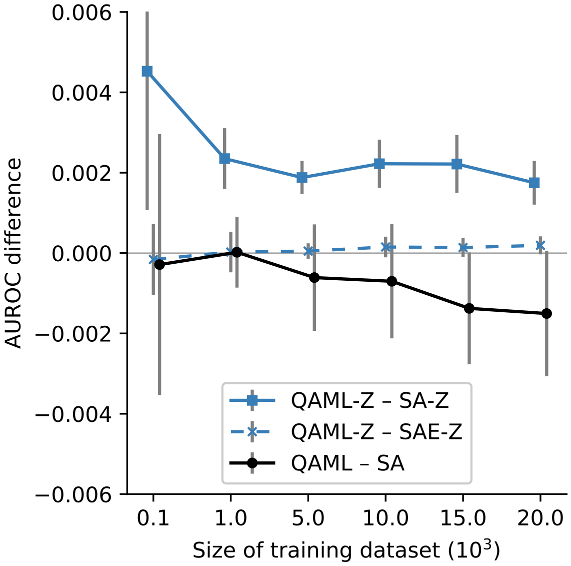

Figure 2: Comparison of quantum annealing and simulated annealing for the new (QAML-Z) and original (QAML) algorithms, measured by area under ROC curve (AUROC). Error bars indicate 1σ error, showing statistically significant improvement of QAML-Z over SA-Z in comparison to the original algorithm and its simulated annealing counterpart. Excited states are included in the SAE-Z variant, reproducing QAML-Z performance.

Since the original simulated annealing benchmark did not ensemble excited states, we report on SA-Z without excited states. However, to attempt to match the improved quantum annealing performance, we also propose SAE-Z, in which the supremum over a set of excited states is used to improve the area under ROC curve (described in Section 3 of the Supplementary Material) in the same manner as for QAML-Z and QAML. We find that SAE-Z performs statistically indistinguishably from QAML-Z, suggesting that excited states can be effectively used to generate stronger classifiers.

4 Conclusion

We find that an extension of QAML to the continuous space over a set of augmented weak classifiers yields strong classifiers that are more competitive with state-of-the-art classical techniques than previously reported.[1] Although QAML-Z remains at a disadvantage to a deep neural network (DNN) for sufficiently large training sets, the performance gap between QAML and DNN has been reduced by a factor of two by applying QAML-Z. Moreover, the QAML advantage over DNN for small training sets has grown to $> 40\%$. The extent of improvement of QAML-Z over DNN for Higgs decay classification suggests that noisy intermediate-scale quantum devices may be approaching real-world applicability in machine learning despite their limitations. Furthermore, the favorable results of zooming in on an Ising model to achieve a solution unreachable by discrete optimization provides future direction for quantum annealing applications, potentially extending to quantum machine learning algorithms beyond QAML.

Acknowledgments

Part of this work was conducted at "*iBanks*," the AI GPU cluster at Caltech. We acknowledge NVIDIA, SuperMicro and the Kavli Foundation for their support of "*iBanks*." This work is partially supported by DOE/HEP QuantISED program grant, Quantum Machine Learning and Quantum Computation Frameworks (QMLQCF) for HEP, award number DE-SC0019227. The work is also supported in part by the AT&T Foundry Innovation Centers through INQNET, a program for accelerating quantum technologies. The work of DL and JJ was partially supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the U.S. Army Research Office contract W911NF-17-C-0050.

References

[1] Alex Mott et al. "Solving a Higgs optimization problem with quantum annealing for machine learning". In: *Nature* 550 (2017), p. 375. URL: http://dx.doi.org/10.1038/nature24047.

[2] Kristen L. Pudenz and Daniel A. Lidar. "Quantum Adiabatic Machine Learning". In: *Quantum Information Processing* 12.5 (May 2013), pp. 2027–2070. ISSN: 1570-0755. DOI: 10.1007/s11128-012-0506-4. URL: http://dx.doi.org/10.1007/s11128-012-0506-4.

[3] P. I. Bunyk et al. "Architectural Considerations in the Design of a Superconducting Quantum Annealing Processor". In: *IEEE Transactions on Applied Superconductivity* 24.4 (2014), pp. 1–10. ISSN: 1051-8223. DOI: 10.1109/TASC.2014.2318294.

[4] Joshua Job and Daniel Lidar. "Test-driving 1000 qubits". In: *Quantum Science and Technology* 3.3 (2018), p. 030501. DOI: 10.1088/2058-9565/aabd9b. URL: https://doi.org/10.1088\%2F2058-9565\%2Faabd9b.

[5] Davide Venturelli et al. "Quantum Optimization of Fully Connected Spin Glasses". In: *Phys. Rev. X* 5 (3 2015), p. 031040. DOI: 10.1103/PhysRevX.5.031040. URL: https://link.aps.org/doi/10.1103/PhysRevX.5.031040.

[6] Vicky Choi. "Minor-embedding in adiabatic quantum computation: I. The parameter setting problem". In: *Quantum Information Processing* 7.5 (2008), pp. 193–209. ISSN: 1573-1332. DOI: 10.1007/s11128-008-0082-9. URL: https://doi.org/10.1007/s11128-008-0082-9.

[7] Vicky Choi. "Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design". In: *Quantum Information Processing* 10.3 (2011), pp. 343–353. ISSN: 1573-1332. DOI: 10.1007/s11128-010-0200-3. URL: https://doi.org/10.1007/s11128-010-0200-3.

[8] Christine Klymko, Blair D. Sullivan, and Travis S. Humble. "Adiabatic quantum programming: minor embedding with hard faults". In: *Quant. Inf. Proc.* 13.3 (2014), pp. 709–729. DOI: 10.1007/s11128-013-0683-9. URL: http://dx.doi.org/10.1007/s11128-013-0683-9.

[9] Jun Cai, William G. Macready, and Aidan Roy. "A practical heuristic for finding graph minors". In: *arXiv:1406.2741* (2014). URL: http://arXiv.org/abs/1406.2741.

Supplementary Material

1 Derivation

To motivate the QAML-Z algorithm and provide the necessary framework for its derivation, we first provide a brief description of the Hamiltonian to minimize the error between a strong classifier and a training set in the original QAML algorithm. Given weak classifiers $c_i(\mathbf{x}_\tau) = \pm 1/N$ and truth label $y_\tau = \pm 1$ over a training set of size S consisting of labeled data $\{\mathbf{x}_\tau, y_\tau\}$, let $R(\mathbf{x}_\tau) = \sum_i^N s_i c_i(\mathbf{x}_\tau)$ be a strong classifier that ensembles the weak classifiers with coeffiicents s_i. To minimize classification error, we simply minimize the distance between y and R:

$$||\mathbf{y} - \mathbf{R}||^2 = \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i c_i(\mathbf{x}_\tau) \right|^2 \tag{10a}$$

$$= ||\mathbf{y}||^2 - 2\sum_{i=1}^{N}\sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) y_\tau + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) s_j c_j(\mathbf{x}_\tau). \tag{10b}$$

Removing the spin-independent term $||\mathbf{y}||^2$ and the self-spin interactions $c_i^2(\mathbf{x}_\tau)$ to create a problem suitable for quantum annealing, we may rewrite the Hamiltonian as follows (scaling by a factor of 2 for convenience):

$$H = \sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) s_j c_j(\mathbf{x}_\tau) - \sum_{i=1}^{N}\sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) y_\tau. \tag{11}$$

To map this Hamiltonian to a continuous space, we wish to find a suitable set of $\{\mu_1, \mu_2, \ldots, \mu_N\}$ where $\mu_i \in [-1, 1]$ replaces each s_i in the above Hamiltonian. By minimizing over $\{\mu_i\}$, we may find a strong classifier given by $R(\mathbf{x}_\tau) = \sum_i^N \mu_i c_i(\mathbf{x}_\tau)$. However, quantum annealing provides the constraint $s_i = \pm 1$, and thus the search space needs to be split in a divide-and-conquer strategy. To probe either end of the search space at each iteration $t = 0, 1, \ldots, T-1$, we update μ_i by centering it around the previous mean and shifting it by the latest annealing result, yielding $\mu_i(t+1) = \mu_i(t) + s_i\sigma(t+1)$, where $\sigma(t) = b^t$ is the search breadth and $0 < b < 1$ is a free parameter. To provide intuition for this update rule, we consider a binary search with one qubit using this scheme, setting $\mu_0(0) = 0$ and $\sigma(t) = 1/2^t$. If, for instance, we find that $s_0 = -1$ at the first annealing, we wish to update $\mu_0(1) = -\frac{1}{2}$ and $\sigma(1) = \frac{1}{2}$. If we then receive $s_0 = +1$ at the second annealing, we wish to update $\mu_0(2) = -\frac{1}{4}$ and $\sigma(2) = \frac{1}{4}$. The weight given to each classifier must also be centered and shifted, yielding the substitution $s_i c_i(\mathbf{x}_\tau) \to (\sigma(t)s_i + \mu_i(t))c_i(\mathbf{x}_\tau)$ in the Hamiltonian:

$$H(t) = \sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{\tau=1}^{S} \big(\sigma(t)s_i c_i(\mathbf{x}_\tau) + \mu_i(t)c_i(\mathbf{x}_\tau)\big)\big(\sigma(t)s_j c_j(\mathbf{x}_\tau) + \mu_j(t)c_j(\mathbf{x}_\tau)\big) - \sum_{i=1}^{N}\sum_{\tau=1}^{S} \big(\sigma(t)s_i c_i(\mathbf{x}_\tau) + \mu_i(t)c_i(\mathbf{x}_\tau)\big)y_\tau \tag{12a}$$

$$= \Bigg[\sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{\tau=1}^{S} \sigma^2(t)c_i(\mathbf{x}_\tau)c_j(\mathbf{x}_\tau)s_i s_j + \sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{\tau=1}^{S} \big(\sigma(t)s_i c_i(\mathbf{x}_\tau)\mu_j(t)c_j(\mathbf{x}_\tau) + \sigma(t)s_j c_j(\mathbf{x}_\tau)\mu_i(t)c_i(\mathbf{x}_\tau)\big)$$
$$- \sum_{i=1}^{N}\sum_{\tau=1}^{S}\sigma(t)s_i c_i(\mathbf{x}_\tau)y_\tau\Bigg] + \Bigg[\sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{\tau=1}^{S}\mu_i(t)\mu_j(t)c_i(\mathbf{x}_\tau)c_j(\mathbf{x}_\tau) - \sum_{i=1}^{N}\sum_{\tau=1}^{S}\mu_i(t)c_i(\mathbf{x}_\tau)y_\tau\Bigg] \tag{12b}$$

The second bracketed term is constant, and thus we may remove it from the Hamiltonian. Furthermore, although we dropped quadratic self-spin terms $c_i^2(\mathbf{x}_\tau)$ earlier, we may recover the linear cross-terms that re-appeared after the substitution when $j = i$. Simplifying, we find the following Hamiltonian:

$$H(t) = \sum_{i=1}^{N}\left(-C_i + \sum_{j=1}^{N}\mu_j(t)C_{ij}\right)\sigma(t)s_i + \sum_{i=1}^{N}\sum_{j>i}^{N} C_{ij}\sigma^2(t)s_i s_j, \tag{13}$$

where we have defined:

$$C_{ij} = \sum_{\tau=1}^{S} c_i(\mathbf{x}_\tau)c_j(\mathbf{x}_\tau), \qquad C_i = \sum_{\tau=1}^{S} c_i(\mathbf{x}_\tau)y_\tau. \tag{14}$$

This is the Hamiltonian we encode on the quantum annealer.

2 Quantum Annealing

After augmenting the original set of 36 classifiers, there are 252 fully connected variables in the Hamiltonian. However, due to its Chimera graph architecture and the fact that only 1098 of the 1152 qubits are functional, the USC-based D-Wave 2X only has 33 fully connected logical qubits. Hence, we prune the cross-terms in the Ising Hamiltonian, retaining only the largest 5% of weights. This allows a minor embedding operation [6, 7, 8, 9] in combination with the classical polynomial-time `fix_variables` procedure in the D-Wave API to program the problem on the quantum annealer. Each logical qubit is mapped to a chain of physical ferromagnetically coupled qubits on the D-Wave machine, where the internal coupling of each chain may be set to prevent thermal excitations and other noise from breaking the chain while still ensuring that the Hamiltonian drives the system dynamics [5]. We set the ratio r between coupling within each chain to the largest coupling in the Hamiltonian, monotonically decaying it with iteration number for the first 5 iterations before fixing it at a constant value: $r = \{3.0, 1.0, 0.5, 0.2, 0.1, 0.1, \dots\}$. Moreover, we reduce random errors on the local fields and couplers by randomizing the encoding via sign flips, annealing over g gauges where $g = \{50, 10, 1, 1, \dots\}$ also varies with iteration number. For each gauge, we sample the D-Wave annealing result 200 times and measure the qubit chains with a majority vote, using a coin-toss tie-breaker for even length chains. Additionally, we perform each annealing for 5 μs, having observed minimal variation ($< 0.1\%$) in area under ROC curve for annealing times ranging from 5 μs to 800 μs.

To prevent the zooming algorithm from getting stuck in local minima and to regularize the optimization scheme for small training sets, we flip qubits at random between successive iterations through two processes: flipping qubits when an update $\mu_i(t+1) = \mu_i(t) + s_i\sigma(t+1)$ yields a worse energy than $\mu_i(t)$ and flipping qubits uniformly randomly. In either case, we flip s_i to $-s_i$, zooming into the opposite region than the one selected by annealing. Since the first case suggests that the energy surface is more indifferent to a qubit flip, we flip qubits with a higher probability each iteration ($p_f = \{0.16, 0.08, 0.04, 0.02, 0.01, 0.01, \dots\}$) compared to the uniform qubit flip case ($q_f = \{0.08, 0.04, 0.02, 0.01, 0.005, 0.005, \dots\}$). Although we find that the heuristic of constant qubit flip probabilities provides more stability than using the Metropolis update transition probability $\min(1, \exp(-\beta\Delta E))$, it is quite possible that other update rules we did not explore (e.g., a genetic algorithm or an update rule dependent on training set size) could further improve the QAML-Z algorithm.

To select excited states for the classifier, we place two criteria: a maximum distance d to the lowest-energy state found (i.e., an excited state must have an energy less than $(1-d)E_{\text{ground}}$ for $E_{\text{ground}} < 0$ or less than $(1+d)E_{\text{ground}}$ for $E_{\text{ground}} > 0$), and a maximum total number of excited states n_e to be selected. To prevent an exponential increase in the tree of excited states generated by the zooming algorithm, we also decay the values of d and e by iteration number, setting $d = \{0.08, 0.04, 0.02, 0.01, 0.01, \dots\}$ and $n_e = \{16, 4, 1, 1, \dots\}$. We then take the supremum over the set of excited states' background rejection values for each efficiency in the ROC curve as done in the original QAML algorithm [1], estimating errors by reweighting samples from a Poisson distribution.

3 Simulated Annealing

We perform simulated annealing using the Metropolis update rule, flipping a random spin to construct a trial spin vector $\vec{s}'$ from the spin vector $\vec{s}$. If the energy $H(\vec{s}') < H(\vec{s})$, then the new vector $\vec{s}'$ is accepted with probability 1. However, if $H(\vec{s}') > H(\vec{s})$, the trial vector is accepted with probability $\exp[-\beta(H(\vec{s}') - H(\vec{s}))]$. After randomly selecting a spin to flip N times (where $\vec{s}$ has N spins), a sweep has been completed. The inverse temperature β is stepped with a linear inverse temperature schedule from $\beta_i = 0.1$ to $\beta_f = 5$ over $W = 1000$ sweeps, incrementing the temperature by $\frac{\beta_f - \beta_i}{W}$ after each sweep. This process is repeated 1000 times, and the lowest-energy state is selected in the SA-Z algorithm. To assemble excited states for the SAE-Z benchmark, we perform 5000 sweeps for 5000 reads and select excited states using the same criteria as for quantum annealing, increasing the simulated annealing runtime to ensure that enough high-quality excited states are found to fill the maximum permitted number of excited states $n_e = \{16, 4, 1, 1, \dots\}$.

Temperature schedules reaching β as large as 10 and performing up to 100,000 sweeps per read were found to have no significant impact on the results. Moreover, we observe that QAML-Z and SA-Z reach statistically identical lowest-energy states with 5000 sweeps and 5000 reads. On the training set, the difference in the area under ROC curve between the lowest-energy QAML-Z state and the SA-Z state is at most on the order of 10^{-5} with a standard deviation on the order of 10^{-5} for all training set sizes. This suggests that both annealing methods found similar ground states at the end of the zooming procedure, although they may have taken different paths to the final state due to the insertion of the Metropolis move. Additionally, with these settings for SA, we observe that SA matches or bests QA with regards to minimum observed energy on the training set when they are each supplied identical QUBOs generated during the zooming algorithm (Figure 3). Since QAML-Z and SAE-Z achieve similar performance on the test set after ensembling excited states (Figure 2), we conclude that the zooming methodology is robust to slight differences in annealing performance.

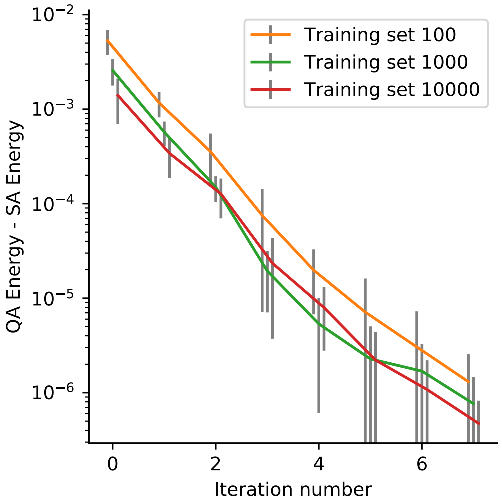

Figure 3: Difference between the lowest energy of quantum annealing (QA) and simulated annealing (SA) on different training sets when given the same QUBO each iteration. SA consistently finds a lower energy on the training set than QA for each given QUBO during zooming. However, we observe no advantage on the test set after excited states are ensembled (Figure 2). Note that although only three training sets are shown here, SA achieves a lower energy than QA for all training set sizes.

4 Zooming and Augmentation Analysis

To provide insight into the impact of iterative zooming, we examine the normalized Ising model energy obtained purely under augmentation but without zooming:

$$H(t) = \frac{1}{S}\sum_{\tau=1}^{S}\left(\sum_{i=1}^{N}\sum_{j>i}^{N}\mu_i(t)c_i(\mathbf{x}_\tau)\mu_j(t)c_j(\mathbf{x}_\tau) - \sum_{i=1}^{N}\mu_i(t)c_i(\mathbf{x}_\tau)y_\tau\right) \tag{15}$$

The results are shown in Figure 4, showing a decrease in energy with iteration number as well as a decreased amount of overfitting with larger training set sizes. However, due to the different number of classifiers, the Ising model energies of augmented and non-augmented classifiers cannot be directly compared. The area under the ROC curve (AUROC) illustrates both the impact of classifier augmentation and the impact of zooming through a direct comparison to QAML (Figure 5), showing advantages in both the classifier augmentation and zooming methodologies.

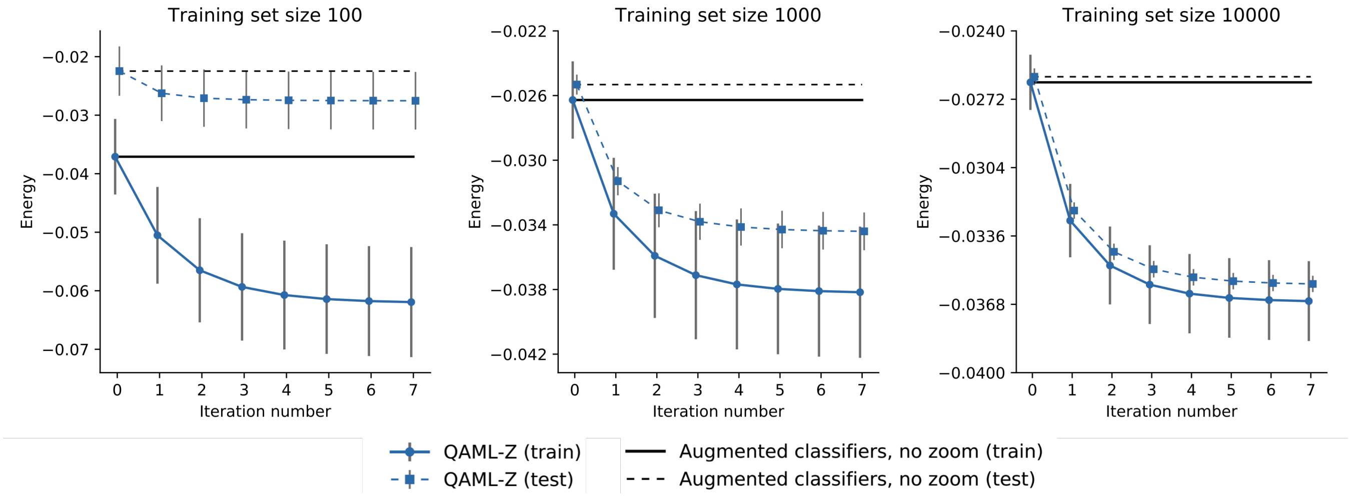

Figure 4: Ising model energy vs. iteration number for different training set sizes. The iterative procedure based on the Hamiltonian (13) monotonically increases AUROC and decreases energy in both the training and test sets. Despite evident overfitting (visible as a large gap between train and test), the performance on the test set steadily improves with iteration number and surpasses the augmented classifiers obtained from Eq. (15). Moreover, overfitting is reduced with larger training set sizes.

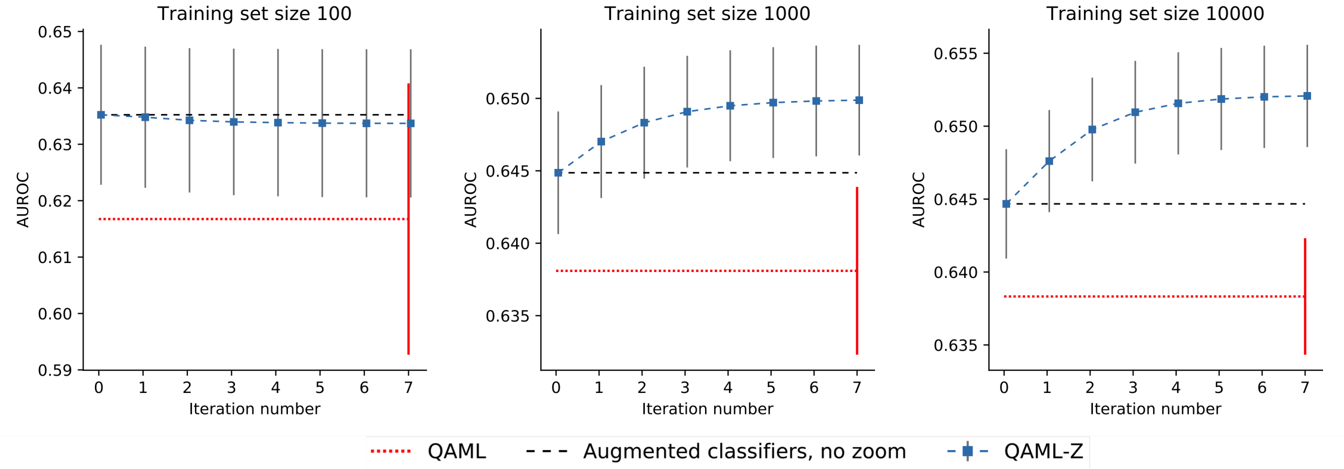

Figure 5: Area under ROC curve (AUROC) vs. iteration number for different training set sizes. Significant improvements by QAML-Z can be separately seen for classifier augmentation (black) and zooming (blue) over the original QAML algorithm (red). The AUROC values shown are determined by the test set. The vertical red lines are the error bars for QAML. Although overfitting occurs in the smallest training size, it may be reduced by limiting iteration number according to cross-validation results.