
Adiabatic Quantum Support Vector Machines

Prasanna Date¹ Dong Jun Woun² Kathleen Hamilton¹ Eduardo A. Coello Perez¹
Mayanka Chandra Shekhar¹ Francisco Rios¹ John Gounley¹ In-Saeng Suh¹ Travis Humble¹
Georgia Tourassi¹

Abstract

Adiabatic quantum computers can solve difficult optimization problems (e.g., the quadratic unconstrained binary optimization problem), and they seem well suited to train machine learning models. In this paper, we describe an adiabatic quantum approach for training support vector machines. We show that the time complexity of our quantum approach is an order of magnitude better than the classical approach. Next, we compare the test accuracy of our quantum approach against a classical approach that uses the Scikit-learn library in Python across five benchmark datasets (Iris, Wisconsin Breast Cancer (WBC), Wine, Digits, and Lambeq). We show that our quantum approach obtains accuracies on par with the classical approach. Finally, we perform a scalability study in which we compute the total training times of the quantum approach and the classical approach with increasing number of features and number of data points in the training dataset. Our scalability results show that the quantum approach obtains a $3.5\text{--}4.5\times$ speedup over the classical approach on datasets with many (millions of) features.

1. Introduction

In the 21st century, computer science has witnessed the rise of machine learning technologies and an explosion of their application areas (LeCun et al., 2015). Machine learning applications now run on all devices—ranging from edge devices such as smartphones to large supercomputing systems (Date, 2019). When developing any end-to-end machine learning application, training the machine learning model takes a significant amount of time and is a major bottleneck (Munson, 2012). Training a machine learning model generally refers to obtaining a set of optimal learning parameters that minimize a well-defined error function (i.e., an optimization problem) (Abu-Mostafa et al., 2012).

Although algorithms for solving optimization problems exist on classical computers, quantum computers are thought to be better at solving them (Moll et al., 2018). Interest

in quantum computing has been growing rapidly in recent years as Moore’s law reaches its inevitable conclusion—a scenario in which classical computing can no longer sustain exponential leaps in performance (Theis & Wong, 2017). For certain problems, some quantum algorithms outperform their classical contemporaries, including the Fourier transformation (Coppersmith, 2002), integer factorization (Shor, 1994), and database search (Grover, 1996), and have attained *quantum supremacy* (Arute et al., 2019). Developing machine learning algorithms for quantum computing should result in shorter training times (Perdomo-Ortiz et al., 2018), but this approach must be tested thoroughly.

To this end, we quantify the gains obtained from using an adiabatic quantum computer (AQC) to train support vector machines (SVM). AQCs are adept at approximately solving the quadratic unconstrained binary optimization (QUBO) problem, which is known to be NP-hard (Kochenberger et al., 2014). In this paper, we formulate SVM training as a QUBO problem and use the D-Wave Advantage AQC to solve it. The main contributions of this work are as follows:

1. We formulate an adiabatic quantum approach for training SVMs.
2. We show that the time complexity of our quantum approach is an order of magnitude faster than the current classical approach.
3. We compare the test accuracies of our quantum approach to those of the classical approach across five benchmark datasets: Iris, Wisconsin Breast Cancer (WBC), Wine, Digits, and Lambeq. Our results show that the quantum approach obtains accuracies on par with the classical approach.
4. We show that the quantum approach is $3.5\text{--}4.5\times$ faster than the classical approach for training SVMs on large (millions of features) synthetic datasets.

2. Related Work

The incorporation of quantum information and quantum computing is expected to have a significant effect on machine learning (Pudenz & Lidar, 2013; Dunjko et al., 2016;

Biamonte et al., 2017). A hallmark of quantum algorithms is leveraging physical phenomena such as superposition and entanglement in addition to an exponentially large Hilbert space for computational advantages (Humble et al., 2022; Delgado et al., 2022). Two main areas of research in quantum machine learning are the development of quantum algorithms that can speed up the training of classical models and the development of quantum models that act as parameterized learning models (Dunjko & Briegel, 2018; Benedetti et al., 2019). Li et al perform a comprehensive review of many quantum classifiers, including support vector machines, and point out challenges in training these models (Li & Deng, 2022).

Quantum machine learning has been explored primarily on universal quantum computers. For example, the Harrow-Hassim-Lloyd algorithm (Harrow et al., 2009) is used to speed up matrix inversion. Jordan’s algorithm (Jordan, 2005) was adapted by Gilyen et al. (Gilyén et al., 2019) to leverage the quantum Fourier transform to compute the gradient of a classical function with a sublinear number of steps. Date proposed the quantum discriminator, which is a quantum discriminant model used for binary classification (Date & Smith, 2024). Quiroga et al. propose the quantum k-means classifier on the IBM quantum computers for discriminating quantum states on hardware (Quiroga et al., 2021).

Quantum gradient descent methods have also seen development (Rebentrost et al., 2019). Furthermore, the general training of machine learning classifiers has been studied, and these investigations show that sublinear training is possible (Li et al., 2019). However, these algorithms require either fault-tolerant quantum systems or quantum registers that far exceed current quantum hardware. Li et al. propose the ST-VQC algorithm to integrate non-linearity in quantum learning and improve the robustness of the learning model to noise. The number of qubits available on near-term, gate-based hardware has limited the direct application of quantum computing to standard machine learning benchmarks. Quantum classifiers that can analyze standard benchmark datasets (e.g., Wine, Sonar) have been studied via numerical simulations (Schuld et al., 2018), but studies that use hardware either require compression methods to reduce the number of training features to a tractable size or are limited to a small number of features (Havlíček et al., 2019). The size of qubit registers continues to increase, and the quality of gates continues to improve; as a result, exploring the applications of parameterized circuit models remains an active area of research (Benedetti et al., 2019).

Instead of implementing a quantum algorithm as a series of unitary gates applied to a qubit register, the D-Wave processor encodes a problem into a system of interacting quantum spins. Through a gradual change in the system’s

Hamiltonian, the process of quantum annealing is used to find the ground state (and correspondingly the solution) of the encoded problem. Casting machine learning tasks as AQC problems through the use of QUBOs is different from constructing a quantum circuit (Pudenz & Lidar, 2013). But, because of the similarities between QUBOs and Ising spin models (Lucas, 2014), several QUBO-based analogues of restricted Boltzmann machines, deep belief networks, and Hopfield networks have been proposed in the literature (Amin et al., 2018; Date et al., 2019b). Chen et al. use the D-Wave system to solve an NP-hard problem that pertains to energy-efficient routing in wireless sensor networks. QUBO-based implementations of conventional machine learning models (e.g., linear regression, k-means clustering, SVMs) have also been developed for the D-Wave platform (Amin et al., 2018; Date et al., 2019b; Arthur & Date, 2021; Date & Potok, 2021; Date et al., 2021). The strength of quantum annealing lies in the ability to solve difficult optimization problems, and this ability means the D-Wave platform could speed up training for classical machine learning models (Neven et al., 2012; Adachi & Henderson, 2015; Willsch et al., 2019).

Recent efforts have explored quantum approaches for SVMs as well. Otgonbaatar successfully demonstrated training SVMs on large remote sensing data by using the D-Wave quantum annealer (Otgonbaatar et al., 2022). To train large datasets, his team employed corsets, which are smaller representative subsets of the data. Barbosa analyzed the factors that contribute to the difficulty of solving a Maximum Clique problem on the D-Wave quantum computer (Barbosa et al., 2021). Lee investigated more effective ways to formulate QUBO problems for linear systems on the D-Wave AQCs (Lee et al., 2022), and Simoes’s research shows that quantum SVMs and neural networks trained on universal quantum computers can achieve higher accuracy than classical approaches (Simões et al., 2023).

3. SVMs

We use the following notation throughout this paper:

- $\mathbb{R}, \mathbb{N}, \mathbb{B}$: Set of real, natural, and binary numbers, respectively.
- X : Training dataset, $X \in \mathbb{R}^{N \times d}$, $N, d \in \mathbb{N}$.
- Y : Labels for binary classification; y_i is $+1$ (-1) if data point $x_i \in X$ belongs to the first (second) class.
- w, b : Weights and bias of the SVM; $w \in \mathbb{R}^d, b \in \mathbb{R}$.

We now state the SVM training problem shown in Figure 1, where we have two classes of data shown by red pluses and blue circles. We would like to find a separating hyperplane,

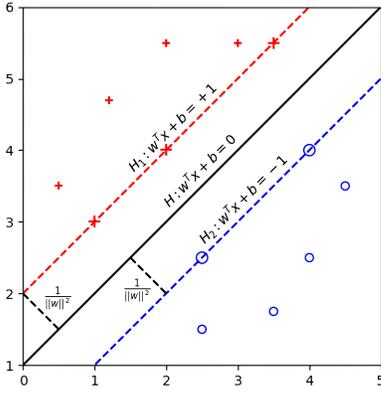


Figure 1: SVMs.

$H : w^T x + b = 0$, that maximizes the distance between the nearest points that belong to the two classes. Hyperplanes $H_1 : w^T x + b = +1$ and $H_2 : w^T x + b = -1$ are parallel to the separating hyperplane H and demarcate the boundary of the corridor in which H lies. H_1 and H_2 can be thought of as the hyperplanes in which the first (H_1) and second (H_2) classes begin. The orthogonal distance between H_1 and H_2 is given by $\frac{2}{\|w\|^2}$, which we would like to maximize. This is equivalent to minimizing $\|w\|^2$.

We must also ensure that all the points are classified correctly. All the red pluses in Figure 1 belong to the first class (i.e., $y_i = +1$). All the blue circles in Figure 1 belong to the second class (i.e., $y_i = -1$). We must ensure that we have $w^T x_i + b \geq 1$ for all the points in the first class and $w^T x_i + b \leq -1$ for all the points in the second class. By incorporating these constraints for all points in the training dataset, the SVM training problem can be stated as follows:

$$\min_{w,b} \|w\|^2 \quad (1)$$

$$\text{subject to: } y_i(w^T x_i + b) \geq 1 \quad \forall i = 1, 2, \dots, N$$

The objective function is convex because its Hessian is positive semi-definite. Furthermore, the constraints are linear, and hence, convex. Problem 1 is therefore a quadratic programming problem. To solve Problem 1, we first compute the Lagrangian dual as follows:

$$\max_{w,b,\lambda} \mathcal{L}(w, b, \lambda) = \|w\|^2 - \sum_{i=1}^N \lambda_i [y_i(w^T x_i + b) - 1], \quad (2)$$

where, λ is the vector that contains all the Lagrangian multipliers (i.e., $\lambda = [\lambda_1 \lambda_2 \dots \lambda_N]^T$, and $\lambda_i \geq 0 \forall i$). The non-zero Lagrangian multipliers in the final solution correspond

to the support vectors and determine the hyperplanes H_1 and H_2 in Figure 1. The Lagrangian dual problem (Eq. 2) is solved in $\mathcal{O}(N^3)$ time on classical computers by applying the Karush-Kuhn-Tucker (KKT) conditions (Karush, 1939; Kuhn & Tucker, 2014). As part of the KKT conditions, we set the gradient of $\mathcal{L}(w, b, \lambda)$ with respect to w to 0. We also set the partial derivative of $\mathcal{L}(w, b, \lambda)$ with respect to b to zero. Doing so yields the following:

$$\begin{aligned} \nabla_w \mathcal{L}(w, b, \lambda) &= w - \sum_{i=1}^N \lambda_i y_i x_i = 0 \\ \therefore w &= \sum_{i=1}^N \lambda_i y_i x_i \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{\partial \mathcal{L}(w, b, \lambda)}{\partial b} &= - \sum_{i=1}^N \lambda_i y_i = 0 \\ \therefore \sum_{i=1}^N \lambda_i y_i &= 0 \end{aligned} \quad (4)$$

Substituting Eqs. 3 and 4 into Eq. 2, we have

$$\mathcal{L}(\lambda) = \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j x_i x_j y_i y_j \quad (5)$$

Note that Eq. 5 is a function of λ only. We want to maximize Eq. 5 for the Lagrangian multipliers and ensure that $\lambda_i, \lambda_j \geq 0 \forall i, j$ while satisfying Eq. 4.

4. Formulation for AQCs

AQCs are adept at solving QUBO problems, which are NP-hard and defined as follows:

$$\min_{z \in \mathbb{B}^M} z^T A z + z^T b, \quad (6)$$

where $z \in \mathbb{B}^M$ is the binary decision vector ($M \in \mathbb{N}$), $A \in \mathbb{R}^{M \times M}$ is the QUBO matrix, and $b \in \mathbb{R}^M$ is the QUBO vector.

To convert the SVM training problem into a QUBO problem, we write Eq. 5 as a minimization problem:

$$\begin{aligned} \min_{\lambda} \mathcal{L}(\lambda) &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j x_i x_j y_i y_j - \sum_{i=1}^N \lambda_i \\ \lambda_i, \lambda_j &\geq 0 \quad \forall i, j \end{aligned} \quad (7)$$

This can be written in matrix form as follows:

$$\min_{\lambda} \mathcal{L}(\lambda) = \frac{1}{2} \lambda^T (X X^T \odot Y Y^T) \lambda - \lambda^T \mathbf{1}_N \quad \lambda \geq 0_N, \quad (8)$$

where 1_N and 0_N represent N -dimensional vectors of ones and zeros, respectively, and \odot is the element-wise multiplication operation.

We now introduce a K -dimensional precision vector, $P = [p_1, p_2, \dots, p_K]^T$, where each entry p_k is a power of 2. This is required to impose the non-negativity constraint on λ . The precision vector must also be sorted. For example, a precision vector could be $P = [\frac{1}{4}, \frac{1}{2}, 1, 2, \dots]^T$. We also introduce K binary variables $\hat{\lambda}_{ik}$ for each Lagrangian multiplier such that

$$\lambda_i = \sum_{k=1}^K p_k \hat{\lambda}_{ik} \quad \forall i = 1, 2, \dots, N, \quad (9)$$

where p_k denotes the k^{th} entry in the precision vector P . Next, we vertically stack all binary variables:

$$\hat{\lambda} = [\hat{\lambda}_{11} \dots \hat{\lambda}_{1K} \hat{\lambda}_{21} \dots \hat{\lambda}_{2K} \dots \hat{\lambda}_{N1} \dots \hat{\lambda}_{NK}]^T \quad (10)$$

We now define a precision matrix as follows:

$$\mathcal{P} = I_N \otimes P^T \quad (11)$$

Notice that

$$\lambda = \mathcal{P} \hat{\lambda} \quad (12)$$

Finally, we substitute the value of λ from Eq. 12 into Eq. 8:

$$\min_{\hat{\lambda} \in \mathbb{B}^{NK}} \mathcal{L}(\hat{\lambda}) = \frac{1}{2} \hat{\lambda}^T \mathcal{P}^T (XX^T \odot YY^T) \mathcal{P} \hat{\lambda} - \hat{\lambda}^T \mathcal{P}^T 1_N \quad (13)$$

Equation 13 is identical to Eq. 6 with $z = \hat{\lambda}$, $A = \frac{1}{2} \mathcal{P}^T (XX^T \odot YY^T) \mathcal{P}$, $b = -\mathcal{P}^T 1_N$, and $M = KN$. Hence, we converted the SVM training problem from Eq. 2 into a QUBO problem in Eq. 13, which can then be solved on AQCs.

4.1. Computational Complexity

We begin our theoretical analysis by defining the space complexity for the number of qubits needed to solve the QUBO. The SVM training problem stated in Eq. 8 contains $\mathcal{O}(N)$ variables (λ) and $\mathcal{O}(Nd)$ data (X and Y). The QUBO formulation of the SVM training problem stated in Eq. 13 consists of the same amount of data. However, as part of the QUBO formulation, we introduced K binary variables for each Lagrangian multiplier in the original problem (Eq. 8). So, the total number of variables in Eq. 13 is $\mathcal{O}(KN)$. So, the qubit footprint (or space complexity) of this formulation would be $\mathcal{O}(N^2 K^2)$ after embedding onto the hardware.

The time complexity of classical SVM algorithms is $\mathcal{O}(N^3)$ (Bottou & Lin, 2007). We analyze the time complexity for our approach in three parts. First, the time complexity of converting Problem 1 into a QUBO problem can be inferred from Eqs. 7 and 9 as $\mathcal{O}(N^2 K^2)$. Because we have $\mathcal{O}(NK)$ variables in the QUBO formulation, embedding can be done in $\mathcal{O}(N^2 K^2)$ time by using the embedding algorithm proposed by Date et al. (Date et al., 2019a). Although the theoretical time complexity of quantum annealing used to obtain an exact solution is exponential ($\mathcal{O}(e^{\sqrt{d}})$) (Mukherjee & Chakrabarti, 2015), a more realistic estimate of the running time can be made by using measures such as ST99 and ST99(OPT) (Wang & Jonckheere, 2019), which give the expected number of iterations to reach a certain level of optimality with 99% certainty. Quantum annealing performs well on problems in which the energy barriers between local optima are tall and narrow because such an energy landscape is more conducive to quantum tunneling. To estimate ST99 and ST99(OPT) for our approach, details on specific instances of the SVM problem are required. Estimating ST99 and ST99(OPT) for generic QUBO formulation of the SVM problem is beyond the scope of the present work.

That said, we would like to shed some light on the quantum annealing running times observed in practice. An AQC can accommodate only finite-sized problems—for example, D-Wave 2000Q can accommodate problems with 64 or fewer binary variables that require all-to-all connectivity (Date et al., 2019a). For problems within this range, a constant annealing time and a constant number of repetitions seem to work well in practice. So, the total time to convert and solve a linear regression problem on an AQC would be $\mathcal{O}(N^2 K^2)$.

Note that the qubit footprint $\mathcal{O}(N^2 K^2)$ and time complexity $\mathcal{O}(N^2 K^2)$ assume that K is a variable. If the precision for all parameters ($\hat{\lambda}$) is fixed (e.g., limited to 32-bit or 64-bit precision), then K becomes a constant factor. The resulting qubit footprint would be $\mathcal{O}(N^2)$, and time complexity would also be $\mathcal{O}(N^2)$. This time complexity is an order of magnitude better than the classical algorithm ($\mathcal{O}(N^3)$).

5. Empirical Analysis

5.1. Methodology and Performance Metrics

Our investigation compares the accuracy of the classical SVM implemented in Scikit-learn with the quantum SVM that utilizes the D-Wave Advantage AQC. In addition to the quantum approach, we also consider simulated annealers to compare accuracy. Scikit-learn solves SVMs by using the sequential minimal optimization algorithm, whereas the quantum approach solves SVMs by transforming the problem into a QUBO problem. In our study, we convert Problem 1 to a Lagrangian dual (Problem 8) and then into a

QUBO (Problem 13) for the quantum and simulated annealers. Our research evaluates the performance of the quantum approach, the classical approach, and the simulated annealing approach across two key metrics: (i) accuracy of the trained model and (ii) total compute time. We calculate the accuracy by dividing the number of correct classifications by the total number of samples. For the classical approach, we measure the compute time as the time taken for training. In contrast, the compute time for the quantum approach encompasses the time required for converting the problem into a QUBO problem, embedding the QUBO problem onto the quantum hardware, and performing quantum annealing to solve the problem. The first two times are based on classical computation because the conversion of an SVM problem to a QUBO problem as well as finding an embedding for the QUBO problem are operations performed on classical computers. The last time, i.e., time taken for quantum annealing, is based on quantum computation. Notably, we observed that the simulated annealers failed to find a viable solution, especially when there were many features, so we decided not to examine the computing time scalability for simulated annealers.

5.2. Hardware Configuration

We conducted the implementations of the classical approach, simulated annealing, and preprocessing stage on two different hardware configurations without altering the underlying methodology. The first system, equipped with an AMD Ryzen 4600H 6-core CPU running at 3 GHz and 16 GB of DDR4 RAM running at 2,666 MHz, was used to evaluate the accuracy results described in Section 5.4. The second system, equipped with an Intel Xeon E5-2690v4 14-core CPU running at 2.6 GHz and 64 GB of DDR4 RAM running at 2,400 MHz, was used to evaluate the compute times presented in Section 5.6. We used the two different machines for the accuracy and compute time comparisons because the AMD machine could not support the larger synthetic datasets used in the compute time experiments.

We used the D-Wave Advantage quantum annealer to evaluate the quantum approach; the annealer offers 5,627 functioning qubits. After preprocessing the problem on a classical computer, we employed the `EmbeddingComposite` class from the D-Wave library to embed the problem onto the quantum hardware. We conducted 10 annealing runs for each problem instance to ensure accuracy and selected the solution with the lowest energy sample.

5.3. Datasets

We compare the classical, quantum, and simulated annealing approaches across the following datasets: synthetic datasets, Iris, WBC, Wine, Digits, and Lambeq. When datasets were split for training and testing, the training data

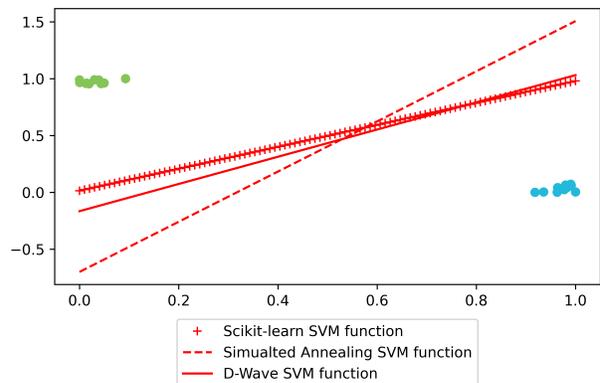


Figure 2: Comparison of hyperplanes created by the support vectors with Scikit-learn (+), simulated annealing(- -), and D-Wave (—) on positive synthetic data (blue and green circles).

was uniformly split per class. For generating the synthetic datasets, we employ two methods. The first method utilizes the `make_blob` function from the Scikit-learn library. This function generates synthetic data with two linearly separable centers and d -dimensional sides. In the second method, we implement a function that generates data points and their corresponding class labels based on the following specified parameters: number of data points, number of features, SVM weights, and SVM bias. We generate each data point by sampling the feature values uniformly at random in the range $[-1, 1]$. We then classify these points based on a linear decision boundary. This process iteratively generates the data until we have the desired number of data points. This approach allows us to generate synthetic datasets that, when trained with an SVM, result in predictable hyperplanes. The Iris dataset contains 150 data points with four features, and the WBC dataset consists of 369 data points with 30 features. The Wine dataset contains 178 data points with 13 features for three classes, and the Digits dataset holds 1,797 data points with 64 features and 10 classes, but we only used two classes of 168 data points.

5.4. Accuracy

We compare the classification accuracy of three distinct SVM approaches: the classical SVM, simulated annealing SVM, and quantum SVM. We employ multiple datasets and conduct 10 rounds of training and testing for each dataset. In many of these datasets, we use only 52 points chosen uniformly at random for training the classifiers because 52 is the largest size of the training dataset that could be accommodated on the annealer. The accuracy results are presented in Table 1.

Table 1: Training and Testing Accuracy of Datasets

Data	Trained Points	Tested Points	Scikit-learn Training Accuracy	Scikit-learn Testing Accuracy	D-Wave Training Accuracy	D-Wave Testing Accuracy	Simulated Annealing Training Accuracy	Simulated Annealing Testing Accuracy
Synthetic positive	20	100	100	100	100	100	100	99.6 ± 0.8
Synthetic negative	20	100	100	100	100	100	100	100
Synthetic random	20	100	100	100	100	100	100	100
Setosa - Virginica	20	80	100	100	100	100	100	99.6 ± 1.2
Setosa - Versicolor	20	80	100	99.6 ± 0.6	100	99.5 ± 1.6	100	99.7 ± 0.5
Versicolor - Virginica	20	80	90.5 ± 6	85.6 ± 3.3	70.5 ± 33.1	67.9 ± 28.8	93.5 ± 6.7	88.1 ± 6.7
WBC	52	517	97.7 ± 1.2	95 ± 1.4	93.3 ± 21	93.1 ± 1.3	91.7 ± 2.6	92.6 ± 1.4
Wine 0—1	52	55	100	100	100	100	99 ± 1.4	99.5 ± 1.7
Wine 0—2	52	78	100	98.2 ± 1.8	96 ± 2.1	96 ± 1.3	95 ± 2.1	94.7 ± 2.3
Wine 1—2	52	67	99.2 ± 1	96 ± 1.7	98.5 ± 1.5	96.6 ± 2.4	96 ± 4	94.2 ± 4.5
Digits 0—1	20	158	100	99.2 ± 1.3	99 ± 2.1	97.5 ± 2.2	—	—
Lambeq	52	30	100	100	93.5 ± 2.9	88.7 ± 4.8	—	—

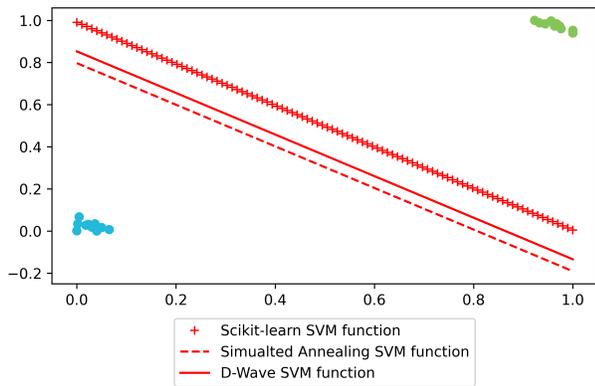


Figure 3: Comparison of hyperplanes created by the support vectors with Scikit-learn (+), simulated annealing(- -) and D-Wave (—) on negative synthetic data (blue and green circles).

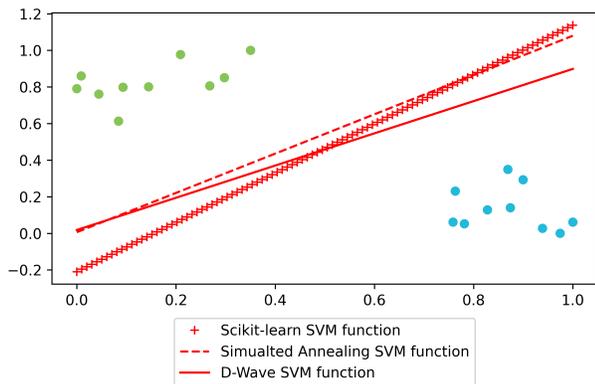


Figure 4: Comparison of hyperplanes created by the support vectors with Scikit-learn (+), simulated annealing(- -), and D-Wave (—) on random synthetic data (blue and green circles split by classification).

5.4.1. SYNTHETIC DATA

We parameterize the synthetic datasets into three types: positive, negative, and random. Positive datasets are designed to be linearly separable when separated by a positive hyperplane (see Figure 2). Conversely, negative datasets are designed to be separated by a negative hyperplane (see Figure 3). To achieve this, we classify randomly sampled points based on a linear decision boundary. Additionally, random datasets with linearly separable data points are generated using the Scikit-learn `make_blob` function (see Figure 4). Before training the SVM models, we ensure that all data points are normalized. Each training dataset consistently contains 20 points with two features. For testing accuracies, we use 100 points. We provide a detailed comparison of scaling data points and features later in this work.

All three approaches—classical SVM, simulated annealing SVM, and quantum SVM—accurately identify the support vectors for the synthetic datasets. Figures 2, 3, and 4 illustrate the datasets and corresponding hyperplanes constructed by the support vectors. Both classical SVM and quantum SVM exhibit 100% accuracy during training and testing across all three datasets. The simulated annealing approach also achieved 100% accuracy on the negative and random datasets. For the positive datasets, the simulated annealing approach achieved an average accuracy of 99.6%. These results highlight the effectiveness of all three methods in accurately identifying support vectors.

5.4.2. IRIS

The Iris dataset consists of 150 samples, three classes, and four features. The three classes describe the Setosa, Versicolor, and Virginica flowers, and the four features represent the length and width of the petal and sepal of the flower. The objective is to perform binary classification for each pair of classes: (Setosa, Versicolor), (Setosa, Virginica), and (Versicolor, Virginica). We normalize the data for training. For each pair, we utilize 20 data points across both classes for training and evaluate the model’s accuracy by using the remaining 80 points. The accuracy results are shown in

Table 2: Training and Testing Accuracy of Varying Annealing Time

Data	Annealing Time(μ s)	Trained Points	Tested Points	Scikit-learn Training Accuracy	Scikit-learn Testing Accuracy	D-Wave Training Accuracy	D-Wave Testing Accuracy	Simulated Annealing Training Accuracy	Simulated Annealing Testing Accuracy
Versicolor - Virginica	20	52	48	93.8 \pm 3.1	94.2 \pm 3.2	87.5 \pm 5.4	86.2 \pm 6.4	87.5 \pm 4.6	85.4 \pm 4.7
Versicolor - Virginica	100	52	48	95.0 \pm 2.3	93.1 \pm 4.3	88.1 \pm 4.1	85.8 \pm 5.8	87.1 \pm 3.4	86.7 \pm 5.0
Versicolor - Virginica	1000	52	48	93.7 \pm 2.4	95.0 \pm 2.2	86.7 \pm 2.8	89.8 \pm 4.0	88.1 \pm 3.2	89.0 \pm 3.9
WBC	20	52	517	95.6 \pm 2.0	95.7 \pm 0.9	90.8 \pm 3.9	92.7 \pm 1.2	91.0 \pm 3.4	93.2 \pm 0.7
WBC	100	52	517	97.3 \pm 2.6	95.4 \pm 1.3	90.6 \pm 3.2	92.9 \pm 1.1	91.3 \pm 2.8	93.2 \pm 1.0
WBC	1000	52	517	96.7 \pm 1.3	95.4 \pm 1.1	91.3 \pm 3.9	92.2 \pm 1.1	91.7 \pm 3.1	92.4 \pm 1.7
Wine 1—2	20	52	67	97.9 \pm 1.9	97.2 \pm 2.0	97.3 \pm 2.7	97.0 \pm 1.4	95.0 \pm 4.3	96.3 \pm 3.3
Wine 1—2	100	52	67	98.7 \pm 1.8	96.3 \pm 1.8	97.5 \pm 2.6	96.0 \pm 2.9	96.7 \pm 3.6	96.7 \pm 2.2
Wine 1—2	1000	52	67	98.8 \pm 1.3	97.2 \pm 2.2	96.7 \pm 2.0	95.7 \pm 2.9	97.1 \pm 1.6	95.7 \pm 4.3

Table 1.

First, all three approaches classify the Setosa and Virginica flowers well. The classical and quantum approaches achieve 100% accuracy in the training and testing data classification. The simulated annealer achieves 100% accuracy in the training data and 99.6% accuracy in the testing data. Second, all three approaches demonstrate near perfect classification performance for Setosa and Versicolor flowers. The classical approach attains an average of 100% training accuracy and 99.6% testing accuracy. The quantum annealer achieves 100% training accuracy and 99.5% testing accuracy. The simulated annealer attains 100% training accuracy and 99.7% testing accuracy. Notably, there is no statistically significant difference in accuracy across the three training methods. Third, all three methods exhibit lower classification performance for the Versicolor and Virginica flower species, which are known to be linearly inseparable. The classical approach achieves 90.5% training accuracy and 85.6% testing accuracy. The quantum annealer achieves 70.5% training accuracy and 67.9% testing accuracy. The simulated annealer achieves 93.5% training accuracy and 88.1% testing accuracy.

5.4.3. WBC

The WBC dataset consists of 569 samples, two classes, and 30 features. The data represents digitalized characteristics of breast cell nuclei. The data is normalized for training. For evaluation, 52 points are used for training, and 517 points are used for testing accuracy. The classical approach demonstrates the highest accuracy and achieves an average of 97.7% for training data and 95.0% for testing data. The quantum approach follows with a training accuracy of 93.3% and a testing accuracy of 93.1%. Last, the simulated annealing approach achieves a training accuracy of 91.7% and a testing accuracy of 92.6%.

5.4.4. WINE

The Wine dataset has 178 samples, three classes, and 13 features. The features describe characteristics of the wine, including alcohol, ash, and magnesium content. We compare the three classes in combination: (class 0 and class 1),

(class 0 and class 2), and (class 1 and class 2). In Table 1, the classes are represented as Wine Class # - Class #. The data is normalized for training, and 52 points are used. The remaining points are used for testing accuracy for the three combinations of datasets.

First, when comparing class 0 and class 1, 52 points are trained and 55 points are tested. The training and testing accuracy averaged 100% for the classical SVM and quantum SVM and 99% and 99.5% for the simulated annealer. Second, when comparing class 0 and class 2, 52 points are trained and 78 points are tested. The training and testing accuracy averaged 100% and 98.2% for the classical SVM. The quantum SVM averaged 96% in accuracy for both training and training, and the simulated annealing SVM averaged 95.0% and 94.7% in accuracy. Third, when comparing class 1 and class 2, 52 points are trained and 67 points are tested. The training and testing accuracy averaged 99.2% and 96.0% for the classical SVM, 98.5% and 96.6% for the quantum SVM, and 96.0% and 94.2% for the simulated annealer.

5.4.5. DIGITS (0–1)

The Digits dataset depicts numerals between 0 and 9 in black-and-white 8×8 images. Each pixel value is between 0 and 16, where 16 is the darkest. Each image has its 64 pixels represented in one row. So, in training, each point has 64 features. The value of each pixel is normalized. We train 20 points each of class 0 and class 1 for 10 iterations to observe the accuracy. To test accuracy, 158 points are used. In Table 1, the classes are represented as Digits Class 0–1. The classical approach resulted in a 100% training accuracy and a 99.2% testing accuracy. The quantum approach resulted in a 99% training accuracy and a 97.5% testing accuracy. The simulated annealer failed to find a solution for the QUBO problems.

5.4.6. LAMBEQ

Lambeq is a dataset consisting of concise three to four word sentences classified into the domains of cooking and technology, with each word labeled according to its respective part of speech (e.g., verb, noun). We train fifty-two of the

Table 3: Scalability with Number of Features

Features	Scikit-learn Time (ms)	D-Wave Embedding Time (ms)	D-Wave Preprocessing Time (ms)	D-Wave Access Time (ms)	D-Wave Compute Time (ms) (Preprocess + Embedding + Access)	D-Wave Network Overhead (ms)
2	1.175 ± 0.072	1,379.469 ± 203.994	239.732 ± 28.529	18.39 ± 0.057	1,637.592 ± 211.0	422,713.522 ± 111,923.977
4	1.158 ± 0.024	1,405.793 ± 377.299	231.087 ± 1.612	18.33 ± 0.095	1,655.21 ± 377.127	330,455.173 ± 90,493.002
8	1.185 ± 0.047	1,164.539 ± 378.807	231.308 ± 1.815	18.29 ± 0.099	1,414.137 ± 378.851	450,387.29 ± 206,827.975
16	1.192 ± 0.014	1,232.807 ± 568.895	232.896 ± 1.886	18.37 ± 0.048	1,484.073 ± 568.916	322,623.635 ± 106,189.784
32	1.18 ± 0.09	1,120.983 ± 97.316	231.457 ± 1.17	18.39 ± 0.057	1,370.829 ± 97.337	370,004.535 ± 126,229.111
64	1.253 ± 0.098	1,381.844 ± 839.621	243.054 ± 30.388	18.38 ± 0.042	1,643.278 ± 836.636	552,217.955 ± 286,747.464
128	1.249 ± 0.043	1,073.689 ± 54.481	234.919 ± 11.097	18.33 ± 0.067	1,326.939 ± 59.699	418,365.701 ± 239,712.534
256	1.206 ± 0.082	1,386.806 ± 161.495	234.472 ± 29.823	18.37 ± 0.048	1,639.648 ± 164.609	310,426.496 ± 138,119.185
512	1.261 ± 0.031	1,502.025 ± 388.979	230.431 ± 2.023	18.35 ± 0.071	1,750.806 ± 388.302	317,514.671 ± 156,945.844
1,024	1.493 ± 0.075	1,423.757 ± 355.441	236.499 ± 16.529	18.29 ± 0.11	1,678.546 ± 353.131	390,337.469 ± 123,918.024
2,048	2.011 ± 0.145	1,646.613 ± 727.704	258.914 ± 42.713	18.34 ± 0.07	1,923.867 ± 720.138	288,349.625 ± 135,981.113
4,096	4.598 ± 0.704	1,482.837 ± 686.698	257.999 ± 40.914	18.34 ± 0.097	1,759.176 ± 681.247	399,891.987 ± 196,877.404
8,192	5.228 ± 0.888	1,280.656 ± 72.575	260.521 ± 40.755	18.39 ± 0.074	1,559.567 ± 105.107	316,780.627 ± 139,171.499
16,384	9.234 ± 0.907	1,321.959 ± 90.673	249.479 ± 35.551	18.37 ± 0.048	1,589.807 ± 112.371	393,245.32 ± 97,614.288
32,768	16.969 ± 1.247	1,577.783 ± 1013.991	242.255 ± 23.858	18.36 ± 0.107	1,838.398 ± 1007.663	347,998.856 ± 129,216.235
65,536	34.06 ± 2.547	1,692.453 ± 1249.103	234.506 ± 3.644	18.37 ± 0.048	1,945.328 ± 1248.834	387,732.537 ± 135,155.523
131,072	76.238 ± 4.305	1,302.487 ± 84.587	254.051 ± 44.079	18.34 ± 0.07	1,574.878 ± 93.238	303,776.797 ± 82,509.468
262,144	182.478 ± 10.178	1,654.28 ± 1011.998	243.556 ± 31.683	18.37 ± 0.062	1,916.393 ± 1013.422	356,133.736 ± 129,222.717
524,288	430.502 ± 26.294	1,530.793 ± 910.643	253.15 ± 42.777	18.37 ± 0.058	1,791.273 ± 908.992	334,643.888 ± 109,913.759
1,048,576	822.135 ± 23.357	1,389.431 ± 115.631	281.621 ± 39.667	18.36 ± 0.07	1,689.412 ± 132.526	404,630.828 ± 274,147.286
2,097,152	1,543.618 ± 139.181	1,318.885 ± 88.77	294.537 ± 42.083	18.35 ± 0.071	1,631.772 ± 107.334	328,081.545 ± 142,239.908
4,194,304	3,166.03 ± 316.12	1,300.346 ± 128.42	294.486 ± 32.706	18.36 ± 0.084	1,613.192 ± 124.671	420,615.812 ± 195,816.586
8,388,608	6,920.407 ± 329.305	1,516.167 ± 332.535	343.214 ± 19.5	18.36 ± 0.052	1,877.741 ± 326.011	428,409.075 ± 256,341.033

Table 4: Scalability with Number of Points

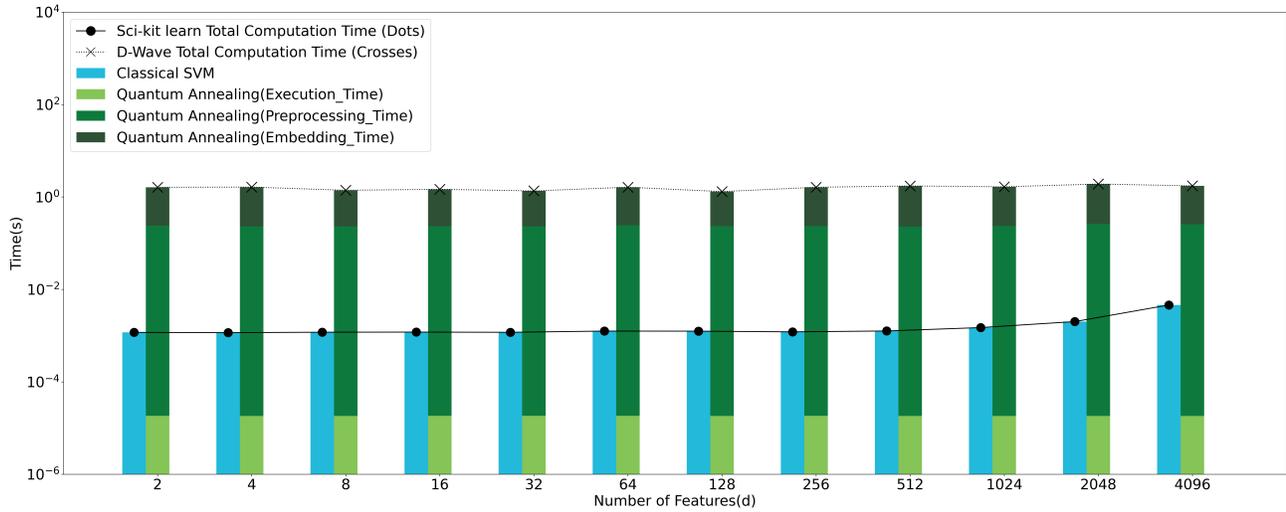
Trained Points	Scikit-learn Time (ms)	D-Wave Embedding Time (ms)	D-Wave Preprocessing Time (ms)	D-Wave Access Time (ms)	D-Wave Compute Time (ms) (Preprocess + Embedding + Access)	D-Wave Network Overhead (ms)
4	397.597 ± 12.209	1,430.915 ± 297.767	52.004 ± 7.233	17.0 ± 0.156	1,499.919 ± 297.346	120.693 ± 15.664
6	628.511 ± 7.604	1,422.757 ± 382.261	71.3 ± 1.359	16.92 ± 0.187	1,510.977 ± 382.923	440.845 ± 81.854
8	865.129 ± 14.27	1,303.998 ± 516.343	83.676 ± 9.532	16.95 ± 0.217	1,404.623 ± 515.02	1,112.126 ± 360.746
10	1,131.034 ± 12.523	1,633.786 ± 985.969	122.318 ± 14.208	17.01 ± 0.311	1,773.115 ± 992.71	2,049.21 ± 661.929
12	1,414.692 ± 42.061	1,334.734 ± 177.382	141.291 ± 11.683	17.03 ± 0.134	1,493.055 ± 179.878	3,905.956 ± 1,651.884
14	1,709.829 ± 30.631	1,378.001 ± 380.399	168.314 ± 2.797	17.29 ± 0.179	1,563.604 ± 382.467	8,745.695 ± 3,737.998
16	2,057.462 ± 21.925	1,154.179 ± 311.386	49.982 ± 2.321	17.18 ± 0.244	1,221.341 ± 310.555	10,334.816 ± 3,908.58
18	2,298.212 ± 102.983	1,217.663 ± 419.797	66.299 ± 11.212	17.1 ± 0.189	1,301.062 ± 419.291	19,508.428 ± 7,737.284
20	2,623.09 ± 145.095	1,257.566 ± 530.37	72.883 ± 11.114	17.32 ± 0.266	1,347.769 ± 527.727	23,161.555 ± 11,342.062
22	3,117.991 ± 116.612	1,370.357 ± 674.162	84.43 ± 13.221	17.47 ± 0.291	1,472.257 ± 686.392	38,135.434 ± 20,395.38
24	3,402.623 ± 140.536	1,225.826 ± 309.729	91.573 ± 4.39	17.62 ± 0.215	1,335.019 ± 308.578	39,167.391 ± 25,129.559
26	3,854.724 ± 161.398	1,078.185 ± 58.742	102.241 ± 3.995	17.66 ± 0.143	1,198.086 ± 59.46	49,917.087 ± 24,788.842
28	4,158.047 ± 237.876	1,409.069 ± 1,002.37	116.227 ± 2.926	17.54 ± 0.272	1,542.835 ± 1,003.154	74,291.927 ± 38,792.915
30	4,325.487 ± 132.322	1,442.387 ± 1,252.714	145.596 ± 30.139	17.85 ± 0.085	1,605.834 ± 1,247.755	84,428.49 ± 44,035.425
32	4,736.955 ± 146.286	1,074.25 ± 54.508	150.959 ± 23.223	17.87 ± 0.048	1,243.08 ± 63.392	109,971.795 ± 33,452.821
34	4,895.237 ± 209.984	1,099.025 ± 75.892	173.559 ± 28.753	17.97 ± 0.095	1,290.554 ± 65.265	100,535.682 ± 52,262.449
36	5,259.747 ± 252.264	2,015.937 ± 2,794.639	183.582 ± 23.139	17.93 ± 0.125	2,217.45 ± 2,791.223	112,656.481 ± 54,607.342
38	5,390.033 ± 195.683	1,079.114 ± 59.131	217.531 ± 35.153	17.98 ± 0.169	1,314.624 ± 62.542	181,915.305 ± 142,277.523
40	5,656.504 ± 101.086	1,868.626 ± 2,396.907	214.248 ± 7.722	18.1 ± 0.067	2,100.973 ± 2,396.668	215,169.1 ± 130,688.927
42	5,962.922 ± 125.236	1,156.912 ± 102.021	237.135 ± 11.716	18.1 ± 0.105	1,412.147 ± 101.156	175,136.911 ± 66,927.252
44	6,253.6 ± 91.927	1,142.343 ± 87.118	246.837 ± 7.396	18.16 ± 0.084	1,407.339 ± 91.611	279,223.122 ± 119,973.036
46	6,463.876 ± 145.813	1,179.851 ± 110.547	291.388 ± 29.706	18.18 ± 0.092	1,489.419 ± 106.889	248,253.607 ± 83,297.36
48	6,771.785 ± 130.662	2,191.238 ± 3,049.933	292.981 ± 12.052	18.31 ± 0.074	2,502.53 ± 3,046.364	297,207.318 ± 178,830.42
50	7,080.465 ± 96.879	1,222.904 ± 114.113	328.075 ± 36.607	18.3 ± 0.125	1,569.279 ± 104.291	292,123.268 ± 141,273.153
52	7,194.556 ± 195.74	1,219.074 ± 88.527	369.37 ± 39.163	18.34 ± 0.07	1,606.784 ± 90.774	407,913.426 ± 209,009.574
54	7,623.851 ± 530.636	4,031.514 ± 8,823.785	383.31 ± 34.857	18.38 ± 0.079	4,433.204 ± 8,824.813	399,734.428 ± 191,104.466

seventy sentences ten times and test thirty points each time. The dataset underwent parsing and embedding by using Bert’s natural language embedder. This was followed by training through an SVM model. The classical approach resulted in a 100% training and testing accuracy. The quantum approach resulted in a 93.5% training accuracy and a 88.7% testing accuracy. The simulated annealer failed to find a solution for the QUBO problems.

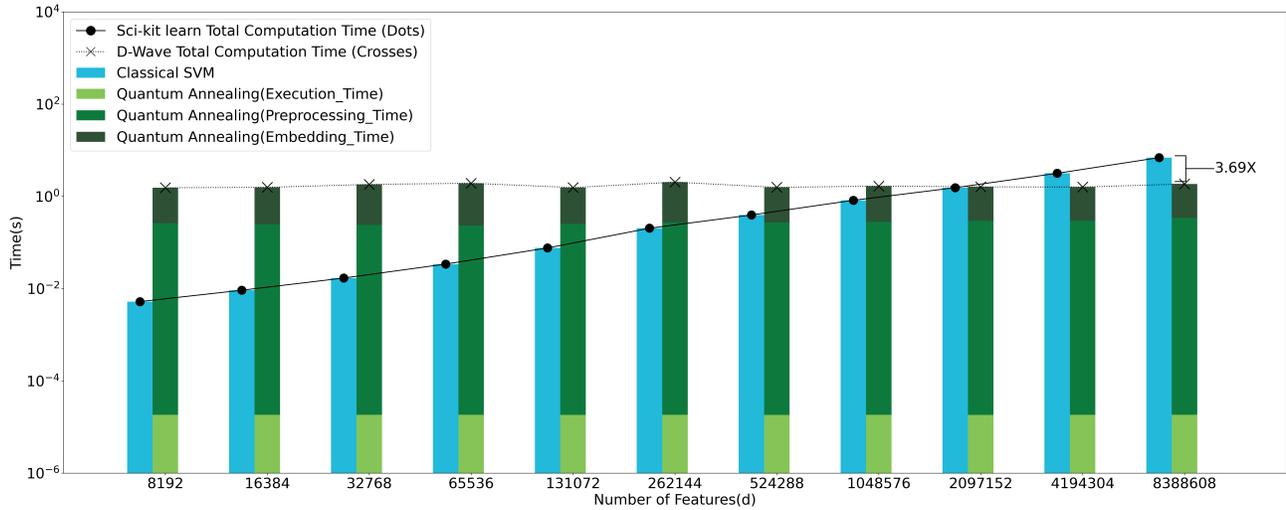
5.5. Annealing Time

The accuracy of the quantum SVM comes second to that of the classical SVM in certain data sets. So, we saw the need to compare the impact of the annealing time parameter, which may influence the quantum annealer’s accuracy. We tested to see if the variance in annealing time has a significant effect on the accuracies of the trained models. Each

Adiabatic Quantum Support Vector Machines



(a) Few features (d)



(b) Many features (d)

Figure 5: Scalability comparison of the Scikit-learn SVM (blue bar and bold line) and quantum SVM (light, medium, and dark green bars and dotted line). The x -axis indicates the number of features (d), and the y -axis logarithmically represents time in seconds. The x -axis ranges from 2 to 8,388,608 across the two figures. In Figure 5a, d varies between 2 and 4,096, and in Figure 5b, d varies between 8,192 and 8,388,608. In Figure 5b, at 8 million features, the quantum approach demonstrated a speedup of $3.69\times$ over the classical approach.

dataset is trained and tested 10 times for each condition. The data used for the experiment included the Iris dataset with the classes Versicolor and Virginica, Wisconsin Breast Cancer, and Wine dataset with the classes 1 and 2. The annealing time used for the experiment included $20\ \mu s$, $100\ \mu s$, and $1000\ \mu s$. Each training set contained 52 training points. The results, which include the mean and standard deviation of the accuracy, are presented in Table 2.

Even as the annealing time increased, the Scikit-learn model consistently performed better and consisted less variance in accuracy than the quantum and simulated annealing mod-

els. There wasn't a statistically significant difference in the accuracy by increasing the annealing time as the mean accuracies of each test were within one standard deviation of each other. However, the Iris data set with the classes Virginica and Versicolor saw less variance and higher training and testing accuracy when trained for $1,000\ \mu s$ as compared to $20\ \mu s$.

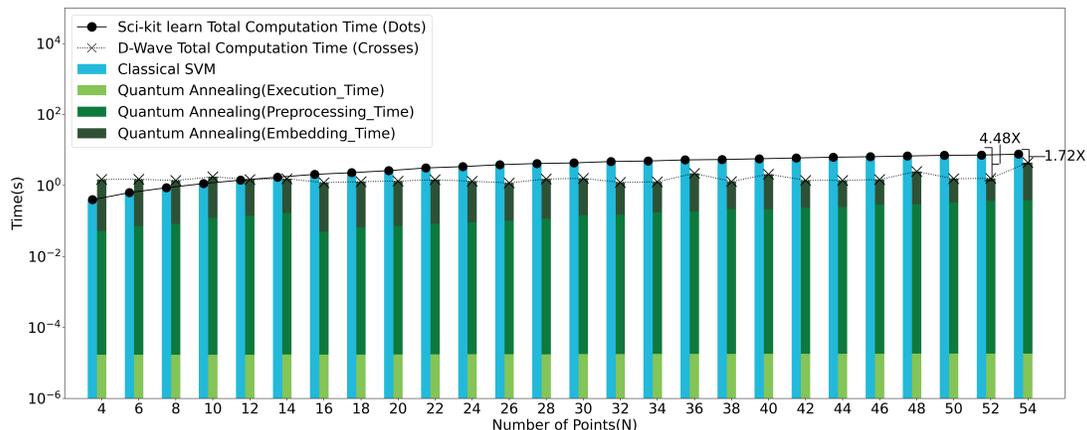


Figure 6: Scalability comparison of the Scikit-learn SVM (blue bar and bold line) and quantum SVM (light, medium, and dark green bars and dotted line). The x -axis indicates the number of points (N), and the y -axis logarithmically represents time in seconds. The x -axis ranges from 4 to 54. In Figure 6, at 52 points, the quantum approach demonstrates a speedup of $4.48\times$ over the classical approach. At 54 points, the quantum approach demonstrates a speedup of $1.72\times$ over the classical approach.

5.6. Time

We conduct a scalability analysis of the training time of the classical and quantum approach using synthetic datasets. We generate the random datasets using the Scikit-learn `make_blob` function. Before training the SVM models, we normalize the data points. The D-Wave access time represents the programming and sampling time. The sampling time includes annealing, read out, delay time. The simulated annealing approach failed to find a solution when there are many features, so we decided it was inappropriate to examine the scalability of the simulated annealing approach. The complexity of the classical approach is $O(N^3)$, and the complexity of the quantum approach is $O(N^2)$.

5.6.1. NUMBER OF FEATURES

We conduct a scalability study to investigate the impact of feature size on the compute time of our classical and quantum SVM approaches. We also describe the runtime performance as the number of features increase from 2 to 8,388,608. The number of data points used for training remains constant at 52 across all the datasets because that is the most data points we could stably embed on the D-Wave Advantage quantum computer. The number of qubits available limits the number of data points. We are also testing each configuration 10 times.

We were able to accommodate SVM problems having millions of features on the D-Wave Advantage system having 5,627 qubits for the following reasons. To begin with, it is clear that the D-Wave hardware has limited qubits and

therefore, strictly constrains the size of the QUBO problem that it can solve. From Eq. 13, we see that the size of the QUBO problem stemming from a given SVM problem solely depends on the number of data points (N) in the SVM problem and not the number of features (d). Therefore, as long as we confine the number of data points, we can keep increasing the number of features indefinitely and still be able to solve the SVM problem on the AQC. In our experiments, we observed that working with around 52 data points resulted in QUBO problems that could be accommodated on the quantum computer. For these problems, an embedding could be found in a reasonable amount of time. Note that the conversion of the SVM problem to the QUBO problem as well as finding an embedding for the QUBO problem happens classically. Once the embedding is found, the annealing is performed on the quantum computer.

The results, including the mean and standard deviation obtained from 10 iterations, are presented in Table 3. The scalability is shown in Figure 5, where the x -axis indicates the number of features (d), and the y -axis logarithmically represents time in seconds. We depict the total Scikit-learn time with light blue bars. The medium green bars represent the D-Wave preprocessing time. The light green bars represent the D-Wave access time, and the dark green bars show the D-Wave embedding time. As shown, the access time stays constant at 18 ms. With smaller number of features ($d \leq 1,048,576$), the classical approach's runtime is faster than the D-Wave approach. As the number of features increases ($d \geq 2,098,152$), the quantum approach's compute time is faster than the classical approach. With

fewer features, the time taken for embedding greatly influences the overall compute time on the D-Wave system. In contrast, the difference in access and preprocessing time has less influence. Similarly, with many features, the embedding time remains the dominant factor in the D-Wave’s compute time, with the access and preprocessing times have minimal impact. However, the preprocessing time does increase, whereas the embedding and access time do not. The performance of both approaches is comparable when there are 2 million features. When training with up to 8 million features, the quantum approach exhibits a speedup of $3.69\times$ over the classical approach. Although not considered for the total computation time, the network overhead time increases as the number of points increases. The increased points result in an increased number of qubits used, which means longer data transfer times to the D-Wave server. The testing and training accuracies are both 100%.

5.6.2. NUMBER OF DATA POINTS

We also conduct a scalability study to investigate how the number of data points affects the compute time of our classical and quantum SVM approaches. We examine the runtime performance as the dataset expands from 4 to 54 data points. The number of features used for training remains constant at 8 million across all the datasets because the runtime performance of both approaches is comparable at 12 points and 8 million features. The number of qubits available limits the number of points.

The results, including the mean and standard deviation obtained from 10 iterations, are presented in Table 4. We present the scalability results in Figure 6, where the x -axis indicates the number of data points (N), and the y -axis logarithmically represents time in seconds. We depict the total Scikit-learn time with blue bars. The green bars represent the D-Wave preprocessing time. The light green bars represent the D-Wave access time, and the dark green bars show the D-Wave embedding time. The dominant factor in the total time taken for the Quantum SVM is the embedding time. The preprocessing and access time increase as the number of points increases, while the embedding time remains relatively constant, except when training 54 points. Considering the available number of qubits, this outcome is expected as the data size increases. However, even at 54 points, the quantum SVM is $1.72\times$ faster than its classical counterpart. At 52 points, where the quantum computer could find an embedding with minimal time fluctuation, the quantum computer outperforms the classical computer by a factor of 4.48.

5.7. Discussion

To the best of our knowledge, this is one of the first studies that shows a demonstrable quantum speedup for training

machine learning models on quantum computers. Our study, coupled with the linear regression study by Date and Potok in 2021 (Date & Potok, 2021), provides a compelling evidence that a quantum speedup can be realized on datasets that have either a large number of data points or a large number of features. On smaller datasets, the overheads of running jobs on the quantum computer overshadow any speedup obtained by the quantum approach.

The testing accuracies presented in Table 1 show that the quantum approach performs better than the classical approach in case of Wine 1–2. On some datasets such as synthetic and Iris, it equals the performance of the classical approach. On other datasets however (Iris Versicolor-Virginica, WBC, Wine 0–2, Digits, and Lambeq), it is outperformed by the classical approach. On these datasets, the lower accuracy of the quantum approach is partly attributed to the fact that the datasets were linearly inseparable, for e.g., Versicolor-Virginica. For other datasets, the accuracy of the weights learned in the quantum approach was directly affected by the numerical precision (i.e., number of bits) allocated to each Lagrangian multiplier in the QUBO problem. This precision was in turn constrained by the hardware architecture, i.e., the number of qubits and their connectivity available on the quantum hardware. Within the constraints imposed by the quantum architecture, the quantum approach was able to produce accuracies that were in the same ballpark as the classical approach for most of the datasets.

Our results in this paper demonstrate that the quantum approach trains SVMs faster than the Scikit-learn classical approach on larger-sized datasets. While the Scikit-learn’s algorithm for training SVMs may not be the most efficient classical algorithm, it certainly is one of the most widely used algorithms in the literature and runs in $\mathcal{O}(N^3)$ time. The classical algorithms have been optimized for decades, whereas the quantum algorithms are still in their nascent stages. In this light, our motivation in this paper was to demonstrate that our quantum approach can perform faster than a widely used classical approach. Our results are intended to serve as a baseline for more optimized quantum approaches in the future. We certainly hope that the future quantum approaches would outperform not only the most widely used classical approaches, but also the *best* performing classical approaches.

6. Conclusion

A critical limitation of current state-of-the-art machine learning is the extensive computational resources required for training. The duration of this process varies significantly—ranging from a few hours to several months—and is contingent upon the size of the training dataset. To address this problem, we introduce a quantum approach for solving the

SVM problem. This paper describes our empirical analysis of transforming SVMs to QUBO problems, which D-Wave Advantage computers can solve. Through theoretical analysis, we establish that our quantum approach outperforms the current classical approach for runtime. Next, we compare the performance of an SVM implementation on a D-Wave Advantage quantum annealer with a classical implementation on AMD Ryzen 5 4600H and Intel Xeon E5-2690v4 CPUs. Our training results demonstrate that the quantum approach's accuracy is comparable with the classical approach. When training with a dataset with 8 million features, the quantum annealer outperforms the classical computer by achieving up to $4.48\times$ faster computation. Our research contribution shows that quantum computing can effectively reduce training times and lead to accelerated scientific discoveries and improved performance for machine learning models.

In the future, we would like to explore novel approaches to train larger datasets effectively. We want to expand our quantum approach to encompass variants of SVMs that leverage kernel methods. Finally, we are interested in exploring techniques to mitigate the effects of noise and errors in quantum annealing.

7. Acknowledgment

This manuscript has been authored by UT-Battelle LLC under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<https://www.energy.gov/doe-public-access-plan>). This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. The authors would like to thank Sam Crawford of Oak Ridge National Laboratory for editing this manuscript.

References

Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. *Learning from data*, volume 4. AMLBook New York, NY, USA:, 2012.

Adachi, S. H. and Henderson, M. P. Application of quantum annealing to training of deep neural networks. *arXiv preprint arXiv:1510.06356*, 2015.

Amin, M. H., Andriyash, E., Rolfe, J., Kulchytskyy, B., and Melko, R. Quantum Boltzmann machine. *Physical Review X*, 8(2):021050, 2018.

Arthur, D. and Date, P. Balanced k-means clustering on an adiabatic quantum computer. *Quantum Information Processing*, 20(9):1–30, 2021.

Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., Biswas, R., Boixo, S., Brandao, F. G., Buell, D. A., et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

Barbosa, A., Pelofske, E., Hahn, G., and Djidjev, H. N. Using machine learning for quantum annealing accuracy prediction. *Algorithms*, 14(6), 2021. ISSN 1999-4893. doi: 10.3390/al14060187. URL <https://www.mdpi.com/1999-4893/14/6/187>.

Benedetti, M., Lloyd, E., Sack, S., and Fiorentini, M. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019.

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., and Lloyd, S. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

Bottou, L. and Lin, C.-J. Support vector machine solvers. *Large scale kernel machines*, 3(1):301–320, 2007.

Coppersmith, D. An approximate fourier transform useful in quantum factoring. *arXiv preprint quant-ph/0201067*, 2002.

Date, P. *Combinatorial Neural Network Training Algorithm for Neuromorphic Computing*. PhD thesis, Rensselaer Polytechnic Institute, 2019.

Date, P. and Potok, T. Adiabatic quantum linear regression. *Scientific reports*, 11(1):21905, 2021.

Date, P. and Smith, W. Quantum discriminator for binary classification. *Scientific Reports*, 14(1):1328, 2024.

Date, P., Patton, R., Schuman, C., and Potok, T. Efficiently embedding QUBO problems on adiabatic quantum computers. *Quantum Information Processing*, 18(4):117, 2019a.

Date, P., Schuman, C., Patton, R., and Potok, T. A classical-quantum hybrid approach for unsupervised probabilistic machine learning. In *Future of Information and Communication Conference*, pp. 98–117. Springer, 2019b.

Date, P., Arthur, D., and Pusey-Nazzaro, L. Qubo formulations for training machine learning models. *Scientific Reports*, 11, 2021.

- Delgado, A., Hamilton, K. E., Vlimant, J.-R., Magano, D., Omar, Y., Bargassa, P., Francis, A., Gianelle, A., Sestini, L., Lucchesi, D., et al. Quantum computing for data analysis in high energy physics. *arXiv preprint arXiv:2203.08805*, 2022.
- Dunjko, V. and Briegel, H. J. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, 81(7):074001, 2018.
- Dunjko, V., Taylor, J. M., and Briegel, H. J. Quantum-enhanced machine learning. *Physical review letters*, 117(13):130501, 2016.
- Gilyén, A., Arunachalam, S., and Wiebe, N. Optimizing quantum optimization algorithms via faster quantum gradient computation. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1425–1444. SIAM, 2019.
- Grover, L. K. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pp. 212–219, 1996.
- Harrow, A. W., Hassidim, A., and Lloyd, S. Quantum algorithm for linear systems of equations. *Physical review letters*, 103(15):150502, 2009.
- Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., and Gambetta, J. M. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.
- Humble, T. S., Delgado, A., Pooser, R., Seck, C., Benink, R., Leyton-Ortega, V., Wang, C.-C. J., Dumitrescu, E., Morris, T., Hamilton, K., et al. Snowmass white paper: Quantum computing systems and software for high-energy physics research. *arXiv preprint arXiv:2203.07091*, 2022.
- Jordan, S. P. Fast quantum algorithm for numerical gradient estimation. *Physical review letters*, 95(5):050501, 2005.
- Karush, W. Minima of functions of several variables with inequalities as side constraints. *M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago*, 1939.
- Kochenberger, G., Hao, J.-K., Glover, F., Lewis, M., Lü, Z., Wang, H., and Wang, Y. The unconstrained binary quadratic programming problem: a survey. *Journal of Combinatorial Optimization*, 28(1):58–81, 2014.
- Kuhn, H. W. and Tucker, A. W. Nonlinear programming. In *Traces and emergence of nonlinear programming*, pp. 247–258. Springer, 2014.
- LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature*, 521(7553):436–444, 2015.
- Lee, H., Noh, S., and Jun, K. Effective QUBO modeling for solving linear systems on D-wave quantum annealing device. In Donkor, E., Hayduk, M., Frey, M. R., Jr., S. J. L., and Myers, J. M. (eds.), *Quantum Information Science, Sensing, and Computation XIV*, volume 12093, pp. 120930F. International Society for Optics and Photonics, SPIE, 2022. doi: 10.1117/12.2632416. URL <https://doi.org/10.1117/12.2632416>.
- Li, T., Chakrabarti, S., and Wu, X. Sublinear quantum algorithms for training linear and kernel-based classifiers. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3815–3824, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL <http://proceedings.mlr.press/v97/li19b.html>.
- Li, W. and Deng, D.-L. Recent advances for quantum classifiers. *Science China Physics, Mechanics & Astronomy*, 65(2):220301, 2022.
- Lucas, A. Ising formulations of many np problems. *Frontiers in Physics*, 2:5, 2014.
- Moll, N., Barkoutsos, P., Bishop, L. S., Chow, J. M., Cross, A., Egger, D. J., Filipp, S., Fuhrer, A., Gambetta, J. M., Ganzhorn, M., et al. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology*, 3(3):030503, 2018.
- Mukherjee, S. and Chakrabarti, B. K. Multivariable optimization: Quantum annealing and computation. *The European Physical Journal Special Topics*, 224(1):17–24, 2015.
- Munson, M. A. A study on the importance of and time spent on different modeling steps. *ACM SIGKDD Explorations Newsletter*, 13(2):65–71, 2012.
- Neven, H., Denchev, V. S., Rose, G., and Mcready, W. G. QBoost: Large scale classifier training with adiabatic quantum optimization. In Hoi, S. C. H. and Buntine, W. (eds.), *Proceedings of the Asian Conference on Machine Learning*, volume 25 of *Proceedings of Machine Learning Research*, pp. 333–348, Singapore Management University, Singapore, 04–06 Nov 2012. PMLR. URL <http://proceedings.mlr.press/v25/neven12.html>.
- Otgonbaatar, S., Datcu, M., and Demir, B. Coreset of hyperspectral images on a small quantum computer. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4923–4926, 2022. doi: 10.1109/IGARSS46834.2022.9884273.

- Perdomo-Ortiz, A., Benedetti, M., Realpe-Gómez, J., and Biswas, R. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. *Quantum Science and Technology*, 3(3):030502, 2018.
- Pudenz, K. L. and Lidar, D. A. Quantum adiabatic machine learning. *Quantum information processing*, 12(5):2027–2070, 2013.
- Quiroga, D., Date, P., and Pooser, R. Discriminating quantum states with quantum machine learning. In *2021 International Conference on Rebooting Computing (ICRC)*, pp. 56–63. IEEE, 2021.
- Rebentrost, P., Schuld, M., Wossnig, L., Petruccione, F., and Lloyd, S. Quantum gradient descent and newton’s method for constrained polynomial optimization. *New Journal of Physics*, 21(7):073023, 2019.
- Schuld, M., Bocharov, A., Svore, K., and Wiebe, N. Circuit-centric quantum classifiers. *arXiv preprint arXiv:1804.00633*, 2018.
- Shor, P. W. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pp. 124–134. Ieee, 1994.
- Simões, R. D. M., Huber, P., Meier, N., Smailov, N., Fuchslin, R. M., and Stockinger, K. Experimental evaluation of quantum machine learning algorithms. *IEEE Access*, 11:6197–6208, 2023. doi: 10.1109/ACCESS.2023.3236409.
- Theis, T. N. and Wong, H.-S. P. The end of moore’s law: A new beginning for information technology. *Computing in Science & Engineering*, 19(2):41–50, 2017.
- Wang, C. and Jonckheere, E. Simulated versus reduced noise quantum annealing in maximum independent set solution to wireless network scheduling. *Quantum Information Processing*, 18(1):1–25, 2019.
- Willsch, D., Willsch, M., De Raedt, H., and Michielsen, K. Support vector machines on the D-Wave quantum annealer. *Computer Physics Communications*, pp. 107006, 2019.