# Adiabatic Quantum Computing for Kernel $k = 2$ Means Clustering

Christian Bauckhage, Cesar Ojeda, Rafet Sifa, and Stefan Wrobel

Fraunhofer Center for Machine Learning, Sankt Augustin, Germany
Fraunhofer IAIS, Sankt Augustin, Germany
B-IT, University of Bonn, Bonn, Germany

**Abstract.** Adiabatic quantum computers are tailored towards finding minimum energy states of Ising models. The quest for implementations of machine learning algorithms on such devices thus is the quest for Ising model (re-)formulations of their underlying objective functions. In this paper, we discuss how to accomplish this for the problem of kernel binary clustering. We then discuss how our models can be solved on an adiabatic quantum computing device. Finally, in simulation experiments, we numerically solve the respective Schrödinger equations and observe our approaches to yield convincing results.

## 1 Introduction

Quantum computing exploits quantum mechanical phenomena for information processing and is now becoming practical. Working quantum computers are on the market [1], industry invests increasing efforts [2,3,4,5,6,7], and further rapid progress is expected [8]. This will likely impact artificial intelligence and machine learning because quantum computing promises efficient solutions to many of the search- or optimization problems encountered in these fields [9,10,11,12,13,14].

Here, we extend our earlier work on quantum computing for unsupervised learning [15] towards the problem of kernel $k = 2$ means clustering and discuss how to solve it via adiabatic quantum computing.

Note that adiabatic quantum computers solve a kind of optimization problem not unfamiliar in machine learning. Devices such as those produced by D-Wave Systems [1,16] determine low energy states of Ising models. While they were originally conceived to describe spin glass systems [17], Ising models occur in other settings, too. Examples include Boolean satisfiability- or graph cutting problems [18,19,20] as well as neurocomputing models known as Hopfield networks [21].

If a problem can be formulated as an Ising energy minimization problem, there are standard procedures for preparing systems of quantum bits (qubits) and energy operators (Hamiltonians) for processing [18,19]. The solution process itself relies on the adiabatic theorem [22] which states that if a quantum system starts in a low energy configuration (ground state) of a Hamiltonian which then gradually changes, the system will end up in the ground state of the resulting Hamiltonian. To harness this for problem solving, one prepares a qubit system

in the ground state of a problem independent Hamiltonian and adiabatically evolves it to a Hamiltonian whose ground state represents a solution to the problem at hand.

In this paper, we discuss these ideas in detail. First, we elaborate on the notion of Ising models and their role in adiabatic quantum computing. We then propose two Ising models for kernel $k = 2$ means clustering. Given these models, we discuss how to set them up for computing and review the required quantum mechanical concepts. Finally, we present several simulation experiments in which we numerically solve the Schrödinger equations which govern the corresponding quantum mechanical processes; these experiments demonstrate the feasibility of our approach and illustrate how appropriately prepared systems of qubits evolve towards a clustering solution.

## 2 Ising Models

Existing adiabatic quantum computers are designed to find low energy states of Ising models. In other words, they solve

$$\boldsymbol{s}^* = \underset{\boldsymbol{s}\in\{-1,+1\}^n}{\operatorname{argmin}} \; \boldsymbol{s}^\intercal \boldsymbol{Q}\boldsymbol{s} + \boldsymbol{s}^\intercal \boldsymbol{q} \tag{1}$$

where the $2^n$ vectors $\boldsymbol{s}$ are possible global states of a system of $n$ entities each of which can be in one of two local states ($+1$ or $-1$). The coupling matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ models interactions within the system and the vector $\boldsymbol{q} \in \mathbb{R}^n$ models external influences.

Since Ising models are concerned with bipolar state vectors $\boldsymbol{s} \in \{-1, +1\}^n$, they appear suited to formalize bi-partitioning problems such as binary clustering of $n$ data points. This is because, for suitable, problem dependent choices of $\boldsymbol{Q}$ and $\boldsymbol{q}$, the entries $s_i^* = \pm 1$ of the solution to (1) can be thought of as membership indicators for two distinct clusters. In section 3, we therefore devise Ising models particularly for this purpose.

Note, however, that the problem in (1) is a quadratic unconstrained binary optimization problem (QUBO) and therefore generally NP hard. For instance, a naïve approach to $k = 2$ means clustering would be to exhaustively evaluate (1) for each of the $2^n$ possible assignments of $n$ data points to 2 clusters. For large $n$, this becomes of course impractical on a digital computer. On an adiabatic quantum computer, on the other hand, we could prepare a system of $n$ qubits that is in a quantum mechanical superposition of all the $2^n$ possible solutions. Here, the challenge is thus to manipulate the system to evolve towards a state that corresponds to a desired partition. In section 4, we discuss how to accomplish this.

## 3 Ising Models for Kernel $k = 2$ Means Clustering

In this section, we devise Ising models for the problem of clustering a sample $X$ of $n$ data points $\boldsymbol{x} \in \mathbb{R}^m$ into two disjoint clusters $X_1$ and $X_2$ where $|X_1| = n_1$,

$|X_2| = n_2$, and $n_1 + n_2 = n$. Without loss of generality, we assume that the data in $X$ are normalized to zero mean so that we have

$$\boldsymbol{\mu} = \frac{1}{n} \sum_{\boldsymbol{x} \in X} \boldsymbol{x} = \frac{1}{n} \sum_{i=1}^{2} \sum_{\boldsymbol{x} \in X_i} \boldsymbol{x} = \frac{1}{n} \big( n_1 \, \boldsymbol{\mu}_1 + n_2 \, \boldsymbol{\mu}_2 \big) = \mathbf{0}. \tag{2}$$

This implies $n_1 \, \boldsymbol{\mu}_1 = -n_2 \, \boldsymbol{\mu}_2$ which is to say that the two cluster means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ will be of opposite sign.

Regarding the idea of $k = 2$ means clustering, our problem would typically be formalized as having to determine the two minimizers $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ of the *within cluster scatter*

$$S_W = \sum_{i=1}^{2} \sum_{\boldsymbol{x} \in X_i} \big\| \boldsymbol{x} - \boldsymbol{\mu}_i \big\|^2. \tag{3}$$

Indeed, most of the well known $k$-means algorithms such as those of Lloyd [23], Hartigan [24], or MacQueen [25] consider this objective.

However, in this paper, we follow a different route and observe that the problem of minimizing the within cluster scatter is equivalent to the problem of maximizing the *between cluster scatter*

$$S_B = \sum_{i,j=1}^{2} n_i n_j \big\| \boldsymbol{\mu}_i - \boldsymbol{\mu}_j \big\|^2. \tag{4}$$

This actually holds true for any $k \geq 1$ and follows from Fisher's analysis of variance [26,27]. It establishes that the *total scatter* can be written as

$$S_T = \sum_{\boldsymbol{x} \in X} \big\| \boldsymbol{x} - \boldsymbol{\mu} \big\|^2 = S_W + \frac{1}{2\,n} S_B \tag{5}$$

which, since $S_T$ and $n$ are positive constants, implies that any decrease of $S_W$ entails an increase of $S_B$.

Looking at the two equivalent objective functions in (3) and (4), we remark that their optimization proves to be NP hard in general [28] because they both constitute integer programming problems in disguise [29]. Algorithms such as those in [23,24,25] are therefore mere heuristics for which there is no guarantee that they will find the optimal solution. In this sense it appears acceptable, that the Ising models we next derive from (4) involve a heuristic assumption, too.

### 3.1 An Ambiguous Ising Model

For our problem of $k = 2$ means clustering, the maximization objective in (4) is expressed in an overly complicated manner and it is easy to see that it can be simplified to

$$S_B = 2\, n_1 n_2 \, \big\| \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \big\|^2. \tag{6}$$

Interestingly, this simplification now provides an intuition as to why $k$-means clustering is agnostic of cluster shapes and distances and often produces clusters

of about equal size [30]. In order for $S_B$ in (6) to be large, both the distance $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|$ between the two cluster centers and the product $n_1 n_2$ of the two cluster sizes have to be large. However, since the sum $n_1 + n_2 = n$ is fixed, the product of the sizes will be maximal if $n_1 = n_2 = \frac{n}{2}$.

This observation provides us with a heuristic argument for how to rewrite the objective in (6) which in turn will allow us to set up an Ising model.

If we assume that, at a solution, we will likely have $n_1 \approx n_2 \approx \frac{n}{2}$, we may consider the approximation

$$\begin{aligned} 2\, n_1 n_2 \left\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right\|^2 &\approx 2\, \tfrac{n^2}{4} \left\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right\|^2 \\ &= 2 \left\|\tfrac{n}{2}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)\right\|^2 \\ &\approx 2 \left\|n_1\, \boldsymbol{\mu}_1 - n_2\, \boldsymbol{\mu}_2\right\|^2 \end{aligned} \tag{7}$$

which turns the $k = 2$ means clustering problem into the problem of having to solve

$$\boldsymbol{\mu}_1^*,\, \boldsymbol{\mu}_2^* = \underset{\boldsymbol{\mu}_1,\, \boldsymbol{\mu}_2}{\operatorname{argmax}} \left\|n_1\, \boldsymbol{\mu}_1 - n_2\, \boldsymbol{\mu}_2\right\|^2. \tag{8}$$

Next, we observe that the norm in (8) can be expressed in a form that does not explicitly depend on the unknown cluster means $\boldsymbol{\mu}_i$. To this end, we gather the given data in a data matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m \times n}$ and introduce two binary indicator vectors $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \{0, 1\}^n$ which indicate cluster memberships in the sense that entry $l$ of $\boldsymbol{z}_i$ equals 1 if $\boldsymbol{x}_l \in X_i$ and 0 otherwise. This way, we can write $n_1\, \boldsymbol{\mu}_1 = \boldsymbol{X}\boldsymbol{z}_1$ as well as $n_2\, \boldsymbol{\mu}_2 = \boldsymbol{X}\boldsymbol{z}_2$ and therefore

$$\left\|n_1\, \boldsymbol{\mu}_1 - n_2\, \boldsymbol{\mu}_2\right\|^2 = \left\|\boldsymbol{X}\left(\boldsymbol{z}_1 - \boldsymbol{z}_2\right)\right\|^2 = \left\|\boldsymbol{X}\boldsymbol{s}\right\|^2. \tag{9}$$

Note that $\boldsymbol{s}$ introduced in (9) is guaranteed to be a bipolar vector because, in $k$-means clustering, every given data point is assigned to one and only one cluster so that

$$\boldsymbol{z}_1 - \boldsymbol{z}_2 = \boldsymbol{z}_1 - (\boldsymbol{1} - \boldsymbol{z}_1) = 2\, \boldsymbol{z}_1 - \boldsymbol{1} = \boldsymbol{s} \in \{-1, 1\}^n. \tag{10}$$

This, however, establishes that there is an Ising model for the problem of $k = 2$ means clustering of zero mean data.

On the one hand, since $\|\boldsymbol{X}\boldsymbol{s}\|^2 = \boldsymbol{s}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{s}$ is convex in $\boldsymbol{s}$, the maximization problem in (8) is equivalent to the following minimization problem

$$\boldsymbol{s}^* = \underset{\boldsymbol{s} \in \{-1,1\}^n}{\operatorname{argmin}}\; -\boldsymbol{s}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\, \boldsymbol{s} \tag{11}$$

$$= \underset{\boldsymbol{s} \in \{-1,1\}^n}{\operatorname{argmin}}\; -\boldsymbol{s}^\intercal \boldsymbol{Q}\, \boldsymbol{s}. \tag{12}$$

On the other hand, because of (2), this will necessarily yield a solution vector $\boldsymbol{s}$ whose entries are not all equal and thus induce a clustering of the data in $\boldsymbol{X}$.

Looking at (12), we next observe that the coupling matrix $\boldsymbol{Q} = \boldsymbol{X}^\intercal \boldsymbol{X}$ is a Gram matrix. The criterion we just derived therefore allows for invoking the

kernel trick and thus leads to kernel $k = 2$ means clustering. In particular, if we choose the coupling matrix for the Ising model in (12) to be a centered kernel matrix

$$Q_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j) - \frac{1}{n}\sum_l k(\boldsymbol{x}_i, \boldsymbol{x}_l) - \frac{1}{n}\sum_k k(\boldsymbol{x}_k, \boldsymbol{x}_j) + \frac{1}{n^2}\sum_{k,l} k(\boldsymbol{x}_k, \boldsymbol{x}_l) \tag{13}$$

where $k\big(\boldsymbol{x}_i, \boldsymbol{x}_j\big)$ is an appropriate Mercer kernel, our assumption of zero mean data remains valid in the feature space as well.

### 3.2 An Unambiguous Ising Model

Looking at (12), we furthermore observe that there is a form of symmetry. This is because, if $\boldsymbol{s}^*$ solves this optimization problem, then so does $-\boldsymbol{s}^*$ since we did not specify whether an entry of, say, $+1$ is supposed to indicate membership to cluster one or two.

To remove this ambiguity, we may remove a degree of freedom from our model. W.l.o.g. we can, for instance, fix $s_n = +1$ and solve (12) for the remaining $n-1$ entries of $\boldsymbol{s}$. This way, the problem becomes to solve

$$\boldsymbol{s}^* = \operatorname*{argmin}_{\boldsymbol{s}\in\{-1,1\}^{n-1}} -\sum_{i,j=1}^{n-1} Q_{ij}\, s_i\, s_j - 2\sum_{j=1}^{n-1} Q_{nj}\, s_j - Q_{nn} \tag{14}$$

$$= \operatorname*{argmin}_{\boldsymbol{s}\in\{-1,1\}^{n-1}} -\boldsymbol{s}^\intercal \boldsymbol{Q}'\, \boldsymbol{s} - \boldsymbol{s}^\intercal \boldsymbol{q}' \tag{15}$$

which we recognize as yet another Ising energy minimization problem.

## 4 Adiabatic Quantum Kernel $k = 2$ Means Clustering

To perform adiabatic quantum kernel $k = 2$ means clustering of $n$ data points, we consider a time-dependent system of $n$ entangled qubits that is in a superposition of $2^n$ basis states. Using the Dirac notation, this is written as

$$\big|\psi(t)\big\rangle = \sum_{i=0}^{2^n-1} a_i(t)\, \big|\psi_i\big\rangle \tag{16}$$

where the time dependent amplitudes $a_i \in \mathbb{C}$ obey $\sum_i |a_i|^2 = 1$. We understand each of the different basis states

$$\big|\psi_0\big\rangle = \big|000\ldots000\big\rangle \tag{17}$$

$$\big|\psi_1\big\rangle = \big|000\ldots001\big\rangle \tag{18}$$

$$\big|\psi_2\big\rangle = \big|000\ldots010\big\rangle \tag{19}$$

$$\big|\psi_3\big\rangle = \big|000\ldots011\big\rangle \tag{20}$$

$$\vdots$$

as an indicator vector which represents one of the $2^n$ possible assignments of $n$ data points to 2 distinct clusters and use the common shorthand to express tensor products, for instance

$$\big|\psi_1\big\rangle = \big|000\ldots001\big\rangle = \big|0\big\rangle \otimes \big|0\big\rangle \otimes \ldots \otimes \big|1\big\rangle. \tag{21}$$

If a quantum system such as the one in (16) evolves under the influence of a time-dependent Hamiltonian $H(t)$, its behavior is governed by the Schrödinger equation

$$\frac{\partial}{\partial t}\big|\psi(t)\big\rangle = -i\,H(t)\,\big|\psi(t)\big\rangle \tag{22}$$

where we have set $\hbar = 1$. In adiabatic quantum computing, we consider periods ranging from $t = 0$ to $t = T$ and assume the Hamiltonian at time $t$ to be given as a convex combination of two static Hamiltonians, namely

$$H(t) = \big(1 - \tfrac{t}{T}\big)\,H_B + \tfrac{t}{T}H_P. \tag{23}$$

$H_B$ is called the *beginning Hamiltonian* whose ground state is easy to construct and $H_P$ is the *problem Hamiltonian* whose ground state encodes the solution to the problem at hand.

For Ising models such as the ones in (12) and (15), there are by now standard suggestions for how to set up a suitable problem Hamiltonian [18,19]. In particular, we may define

$$H_P = \sum_{i,j=1}^{n} Q_{ij}\,\sigma_z^i\,\sigma_z^j \tag{24}$$

where $\sigma_z^i$ denotes the Pauli spin matrix $\sigma_z$ acting on the $i$th qubit, that is

$$\sigma_z^i = \underbrace{I \otimes I \otimes \ldots \otimes I}_{i-1\,\text{terms}} \otimes \sigma_z \otimes \underbrace{I \otimes I \ldots \otimes I}_{n-i\ \text{terms}}. \tag{25}$$

The beginning Hamiltonian is then typically chosen to be orthogonal to the problem Hamiltonian, for instance

$$H_B = -\sum_{i=1}^{n} \sigma_x^i \tag{26}$$

where $\sigma_x^i$ is defined as above, this time with respect to the Pauli spin matrix $\sigma_x$.

To compute a clustering, we then let $|\psi(t)\rangle$ evolve from $|\psi(0)\rangle$ to $\big|\psi(T)\big\rangle$ where $|\psi(0)\rangle$ is chosen to be the ground state of $H_B$. That is, if $\lambda$ denotes the smallest eigenvalue of $H_B$, the initial state $|\psi(0)\rangle$ of the system corresponds to the solution of

$$H_B\big|\psi(0)\big\rangle = \lambda\,\big|\psi(0)\big\rangle. \tag{27}$$

Finally, at time $t = T$, a measurement is performed on the $n$ qubit system. This will cause the wave function $\big|\psi(T)\big\rangle$ to collapse to a particular basis state

and the probability for this state to be $\left|\psi_i\right\rangle$ is given by the amplitude $|a_i(T)|^2$. Yet, since the adiabatic evolution was steered towards the problem Hamiltonian $H_P$, basis states that correspond to ground states of $H_P$ are more likely to be found.

On an adiabatic quantum computer, this algorithm is carried out physically. On a digital computer, we may simulate it by numerically solving

$$\left|\psi(T)\right\rangle = -i \int_0^T H(t) \left|\psi(t)\right\rangle dt \tag{28}$$

which is the approach we adhere to in the next section.

## 5 Practical Examples

In this section, we present several examples which demonstrate the feasibility of quantum computing for kernel $k = 2$ means clustering. Our examples are of didactic nature and first and foremost intended to illustrate the adiabatic evolution of $n$ qubit systems. In each experiment, we therefore restrict ourselves to $n = 16$ data points $\boldsymbol{x}_i \in \mathbb{R}^2$ which form two clusters. In other words, we consider data matrices $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2] \in \mathbb{R}^{2\times 16}$ whose 16 column vectors form two clusters of size $n_1$ and $n_2$, respectively. These simple settings allow us to comprehensibly visualize the data and the evolution of the the amplitudes of the $2^{16}$ or $2^{15}$ basis states of qubit systems which implement the ambiguous and unambiguous model derived above.

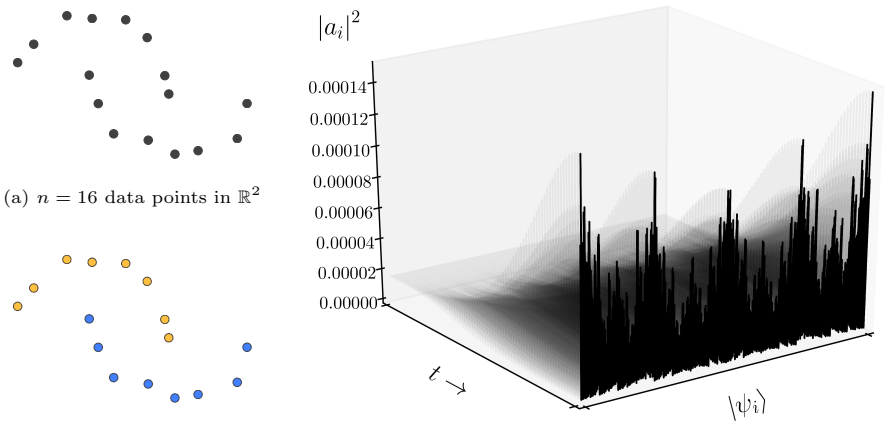
In each experiment, we simulate quantum adiabatic evolutions on a digital computer. To this end, we set up the corresponding problem Hamiltonian $H_P$, the beginning Hamiltonian $H_B$, and its ground state $\left|\psi(0)\right\rangle$ as discussed above and use the *Python* quantum computing toolbox *QuTiP* [31] to numerically solve (28) for $t \in [0, T = 75]$ where $75 \in O(\sqrt{2^n})$.

*Experiment 1:* In our first experiment, we consider the data in Fig. 1(a) and minimize the energy of the ambiguous Ising model in (12) where the coupling matrix $\boldsymbol{Q}$ results from computing (13) with a Gaussian kernel

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{1}{2\,\sigma^2} \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right). \tag{29}$$

Figure 1(c) illustrates the temporal evolution of the amplitudes $|a_i(t)|^2$ of the $2^{16} = 65536$ basis states $\left|\psi_i\right\rangle$ the corresponding 16 qubit quantum system $\left|\psi(t)\right\rangle$ can be in. At $t = 0$, all states are equally likely but over time their amplitudes begin to increase or decrease. At $t = T$, two of the basis states have an amplitude considerably higher than the others so that a measurement will likely cause the system to collapse to either of these equally likely more probable states. These two basis states are $\left|0000000011111111\right\rangle$ and $\left|1111111100000000\right\rangle$ which, when understood as cluster indicator vectors, both induce the result in Fig. 1(b).

Looking at this result, we can conclude that our approach can cluster the data in a manner a human observer would expect and deem appropriate.




(a) $n = 16$ data points in $\mathbb{R}^2$

(b) clustering result

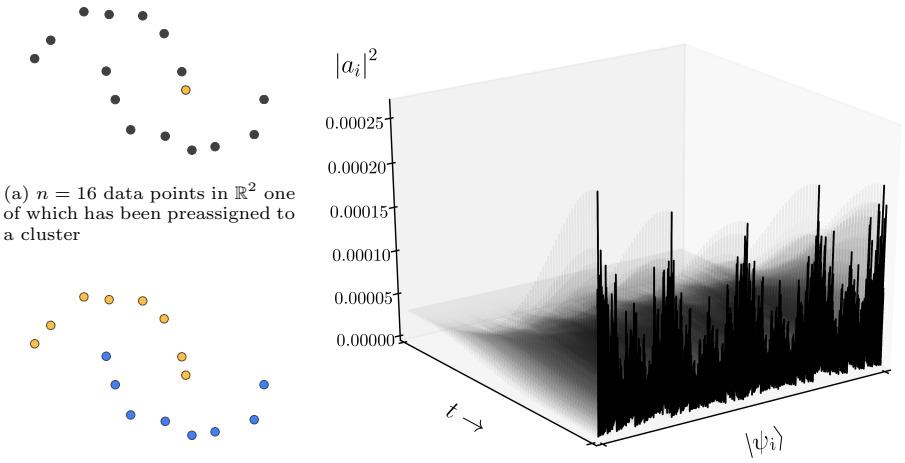
(c) evolution of amplitudes of basis states $|\psi_i\rangle$

Fig. 1: Example of adiabatic quantum kernel $k = 2$ means clustering using the ambiguous Ising model in (12). (a) sample of 16 data points forming two half-moons; (b) corresponding clustering result; (c) adiabatic evolution of a system of 16 qubits. During its evolution over time $t$, the system is in a superposition of $2^{16} = 65536$ basis states $|\psi_i\rangle$ each representing a possible binary clustering. Initially, it is equally likely to find the system in any of these states. At the end, two basis states have noticeably higher amplitudes $|a_i|^2$ than the others and are therefore more likely to be measured; these are $|0000000011111111\rangle$ and $|1111111100000000\rangle$ and they both induce the result in (b).

*Experiment 2:* In our second experiment, we consider the same data as above where, this time, one of the data points has been manually preassigned to a cluster (see Fig. 2(a)). This allows for the use of the unambiguous Ising model in (15). We compute the coupling matrix as above, however, since only 15 data points still need to be assigned to a cluster, we consider a 15 qubit system $|\psi(t)\rangle$ which is in a superposition of $2^{15} = 32768$ basis states.

Figure 2(c) visualizes the adiabatic quantum evolution of this system. Again, at time $t = 0$, each basis state $|\psi_i\rangle$ is equally likely to be measured but the corresponding amplitudes $|a_i|^2$ soon begin to increase or to decrease. Since we are considering the unambiguous Ising model, the process reaches a configuration at $t = T$ where only one basis state has a much higher amplitude than the others. This one is $|000000001111111\rangle$ and induces the result in Fig. 2(b)). Just as in our first experiment, the result obtained from the unambiguous model considered here is reasonable and convincing.

*Experiment 3:* In our third experiment, we investigate whether or not practical success of our approach critically hinges on the heuristic assumption in (7), namely, that clusters are of about equal size. The $n = 16$ data points in Fig. 3(a)




(a) $n = 16$ data points in $\mathbb{R}^2$ one of which has been preassigned to a cluster

(b) clustering result

(c) evolution of amplitudes of basis states $|\psi_i\rangle$

Fig. 2: Example of adiabatic quantum kernel $k = 2$-means clustering using the unambiguous Ising model in (15). (a) sample of 16 data points where one data point has been manually preassigned to a cluster; (b) corresponding clustering result; (c) adiabatic evolution of a system of 15 qubits. Throughout, the system is in a superposition of $2^{15} = 32768$ basis states. Upon termination of its adiabatic evolution, the single most likely basis state for the system to be found in is $|000000001111111\rangle$ which induces the assignment of points to clusters in (b).

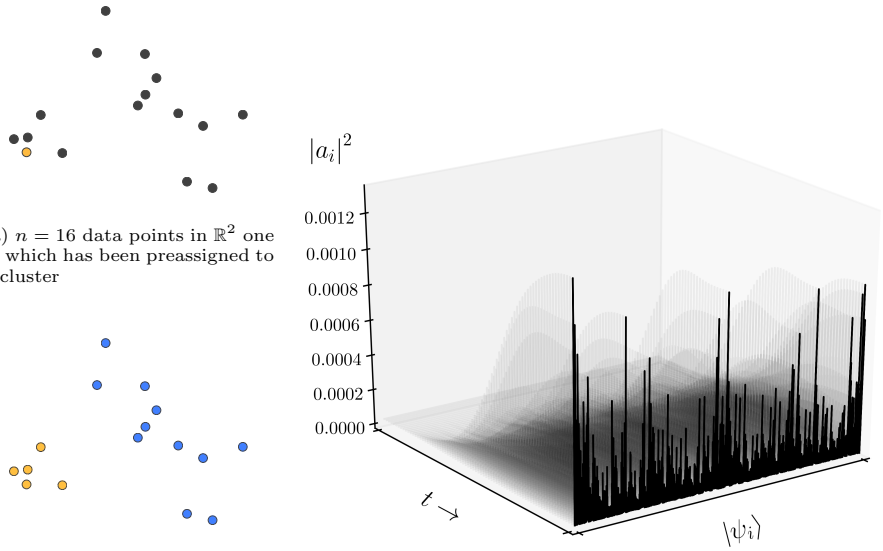
were sampled from two bi-variate Gaussians where $n_1 = 11$ and $n_2 = 5$ and thus form two clusters where the one is more than twice as big than the other.

Since one of the data points in Fig. 3(a) has bee preassigned a cluster label, we again consider the unambiguous Ising model in (15) using a kernelized coupling matrix as described above. The evolution of the corresponding 15 qubit system is shown in Fig. 3(c) and it leads to a configuration where the single most likely basis state $|0000000000001111\rangle$ induces the clustering shown in Fig. 3(b).

As the result in Fig. 3(b) certainly appears reasonable, this experiments shows that the minimum energy configurations of our Ising model(s) for quantum clustering do not necessarily have to correspond to equally sized partitions of a given set of data. This is of course desirable and shows resilience against the simple heuristic we applied in (7). Of course, the result may have looked less convincing to the human eye if the two clusters were closer together; but this caveat would apply to conventional (kernel) $k$-means clustering, too [30].

## 6 Summary

After decades of mainly theoretical research, quantum computing is now about to become practical. Companies such as Google, IBM, Intel, or Microsoft invest




(a) $n = 16$ data points in $\mathbb{R}^2$ one of which has been preassigned to a cluster

(b) clustering result

(c) evolution of amplitudes of basis states $|\psi_i\rangle$

Fig. 3: Example of adiabatic quantum kernel $k = 2$-means clustering applied to a set of data that consists of two clusters of unequal sizes. Nevertheless, using the unambiguous Ising model in (15), the corresponding 15 qubit system evolves to a configuration where the single most likely basis state $|0000000000001111\rangle$ partitions the data into two groups a human observer would deem reasonable.

increasing resources into corresponding research and development and further rapid technological progress is expected. These developments will likely impact supervised and unsupervised machine learning, because working quantum computers promise fast solutions to the kind of search- or optimization procedures that are at the heart of many algorithms in these areas.

In this paper, we were thus concerned with the general feasibility of quantum computing for machine learning and considered adiabatic quantum computing for the problem of kernel $k = 2$ means clustering. We discussed that, from an abstract point of view, the problem of setting up machine learning algorithms for adiabatic quantum computing can be seen as the problem of expressing their objective functions in terms of Ising energy minimization problems because adiabatic quantum computers are tailored towards minimizing Ising energies.

We therefore devised Ising models for (kernel) $k = 2$ means clustering of $n$ data points. The first model was straightforward to derive from an alternative, less well known objective for $k$-means clustering but suffers from ambiguities because if state $\boldsymbol{s}^*$ would minimize the Ising energy, then so would state $-\boldsymbol{s}^*$. We addressed this issue and devised a second, slightly more involved Ising model of $n - 1$ rather than of $n$ degrees of freedom.

In order for this paper to be as self-contained as possible we then discussed how to prepare systems of $n$ or $n-1$ qubits whose adiabatic evolution according to an appropriate time-dependent Hamiltonian would lead to a solution of our Ising energy minimization problems and thus to an assignment of data points to clusters.

Finally, we presented several simulation experiments where we numerically solved the Schrödinger equations governing the dynamics of the corresponding qubit systems. Our examples demonstrated that adiabatic quantum computing can indeed perform kernel $k = 2$ means clustering.

## References


1. D-Wave press release: D-Wave announces D-Wave 2000Q quantum computer and first system order (Jan 2017)
2. Connover, E.: Google Moves toward Quantum Supremacy with 72-qubit Computer. Science News **193**(6) (2018) 13
3. Daimler press release: Daimler joins forces with Google to research the application of quantum computers (Mar 2018)
4. IBM press release: IBM announces advances to IBM quantum systems & ecosystem (Nov 2017)
5. Intel press release: Intel advances quantum and neuromorphic computing research (Jan 2018)
6. Microsoft press release: With new Microsoft breakthroughs, general purpose quantum computing moves closer to reality (2017)
7. VW press release: Volkswagen Group and Google work together on quantum computers (Nov 2017)
8. Knight, W.: Serious quantum computers are finally here. What are we going to do with them? MIT Technology Review (February 2018)
9. Aïmeur, E., Brassard, G., Gambs, S.: Quantum Speed-up for Unsupervised Learning. Machine Learning **90**(2) (2013)
10. Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum Algorithms for Supervised and Unsupervised Machine Learning. arXiv:1307.0411 [quant-ph] (2013)
11. Wittek, P.: Quantum Machine Learning. Academic Press (2014)
12. Schuld, M., Sinayskiy, I., Petruccione, F.: An Introduction to Quantum Machine Learning. Contemporary Physics **56**(2) (2014)
13. Wiebe, N., Kapoor, A., Svore, K.: Quantum Algorithms for Nearest-Neighbor Methods for Supervised and Unsupervised Learning. Quantum Information & Computation **15**(3–4) (2015)
14. Dunjiko, V., Taylor, J., Briegel, H.: Quantum-Enhanced Machine Learning. Physical Review Letters **117**(13) (2016)
15. Bauckhage, C., Brito, E., Cvejoski, K., Ojeda, C., Sifa, R., Wrobel, S.: Ising Models for Biary Clustering via Adiabatic Quantum Computing. In: Proc. EMMCVPR. Volume 10746 of LNCS., Springer (2017)
16. Johnson, M., et al.: Quantum Annealing with Manufactured Spins. Nature **473**(7346) (2011)
17. Ising, E.: Beitrag zur Theorie des Ferromagnetismus. Zeitschrift für Physik **31**(1) (1925) 253–258
18. Farhi, E., Goldstone, J., Gutmann, S., Sipser, M.: Quantum Computation by Adiabatic Evolution. arXiv:quant-ph/0001106 (2000)

19. Lucas, A.: Ising Formulations of Many NP Problems. Frontiers in Physics **2**(5) (2014)
20. Ushijima-Mwesigwa, H., Negre, C., Mniszewski, S.: Graph Partitioning Using Quantum Annealing on the D-Wave System. In: Proc. Int. Workshop on Post Moores Era Supercomputing, ACM (2017)
21. Hopfield, J.: Neural Networks and Physical Systems with Collective Computational Abilities. PNAS **79**(8) (1982)
22. Born, M., Fock, V.: Beweis des Adiabatensatzes. Zeitschrift für Physik **51**(3–4) (1928)
23. Lloyd, S.: Least Squares Quantization in PCM. IEEE Trans. Information Theory **28**(2) (1982)
24. Hartigan, J., Wong, M.: Algorithm AS 136: A $k$-Means Clustering Algorithm. J. of the Royal Statistical Society C **28**(1) (1979)
25. MacQueen, J.: Some Methods for Classification and Analysis of Multivariate Observations. In: Proc. Berkeley Symp. on Mathematical Statistics and Probability. (1967)
26. Fisher, R.: On the Probable Error of a Coefficient Correlation Deduced from a Small Sample. Metron **1** (1921)
27. Bauckhage, C.: $k$-Means and Fisher's Analysis of Variance. Technical report, researchgate (May 2018)
28. Aloise, D., Deshapande, A., Hansen, P., Popat, P.: NP-Hardness of Euclidean Sum-of-Squares Clustering. Machine Learning **75**(2) (2009)
29. Bauckhage, C.: k-Means Clustering via the Frank-Wolfe Algorithm. In: Proceedings KDML-LWA. (2016)
30. MacKay, D.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press (2003)
31. Johansson, J., Nation, P., Nori, F.: QuTiP 2: A Python Framework for the Dynamics of Open Quantum Systems. Computer Physics Communications **184**(4) (2013)