Probabilistic Sampling of Balanced K-Means using Adiabatic Quantum Computing

Jan-Nico Zaech [1,2] Martin Danelljan [1] Tolga Birdal [3] Luc Van Gool [1,2]

[1]ETH Zurich, Switzerland [2] INSAIT, Sofia University, Bulgaria
[3]Imperial College London, United Kingdom

Abstract

Adiabatic quantum computing (AQC) is a promising approach for discrete and often NP-hard optimization problems. Current AQCs allow to implement problems of research interest, which has sparked the development of quantum representations for many computer vision tasks. Despite requiring multiple measurements from the noisy AQC, current approaches only utilize the best measurement, discarding information contained in the remaining ones. In this work, we explore the potential of using this information for probabilistic balanced k-means clustering. Instead of discarding non-optimal solutions, we propose to use them to compute calibrated posterior probabilities with little additional compute cost. This allows us to identify ambiguous solutions and data points, which we demonstrate on a D-Wave AQC on synthetic tasks and real visual data.

1. Introduction

Clustering and uncertainty quantification (UQ) are fundamental problems in machine learning and computer vision. Clustering, the task of grouping objects based on the similarity of their features, plays a pivotal role in analysis and organization of vast quantities of unlabeled visual data with applications including image classification [16, 51, 65], segmentation [23, 32], tracking [48], and network training [22, 75, 76]. UQ, on the other hand, aims to assess the trustworthiness of predictions and accounts for the impact of data variability. As a crucial element for certifying confidence in results, UQ is indispensable in clustering, especially given the often ambiguous nature of data partitioning [41].

Clustering typically involves solving for an optimal assignment of data points to (latent) centroids, while a full Bayesian UQ approach strives to accurately depict the posterior distribution. Unfortunately, both of these problems are known to be notoriously difficult, classified within the NP complexity class[1] [24, 26, 55]. These challenges within the Turingian modes of computation have motivated: (i) continuous relaxations in clustering [78] and (ii) the advent of approximate Bayesian inference methods for UQ, including the EM algorithm [30], variational inference [13], and sequential MCMC [29].

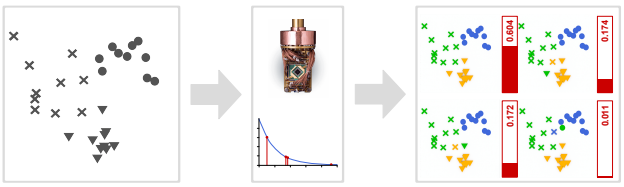


Figure 1. The proposed approach uses an adiabatic quantum computer to sample solutions of a balanced k-means problem. By using an energy-based formulation, likely solutions are drawn from a Boltzmann distribution. By reparametrizing the distribution, the calibrated posterior probability of each solution can be estimated.

In this paper, departing from the above-mentioned approximations or continuous relaxations, we take a completely different approach and harness the upcoming and novel computational paradigm of quantum computers to tackle the challenging task of clustering with *calibrated* uncertainty quantification. Specifically, we focus on the widely used balanced K-means [52, 54, 56], which is an iterative algorithm that first assigns each data point to a cluster followed by an update of cluster centroids.

Despite their early state, quantum machine learning algorithms have approached tasks such as optimization of quadratic problems [47], training of restricted Boltzmann machines [33], and learning with quantum neural networks [1]. While the current quantum computers can only solve small-scale problems, they provide the basis to develop and test algorithms that can considerably increase the size of feasible problems in the future. In our work, we interpret K-means clustering in an energy-based framework

[1]Computing marginal probabilities on a Bayesian network is, basically, a counting problem that can become arbitrarily hard to solve.

and aim to model the posterior distribution by sampling multiple high-probability binary assignments from it on actual quantum hardware at little additional cost. We specifically use the *quantum annealer* (QA) of D-Wave [58], which follows the concept of *adiabatic quantum computing* (AQC) [10] and avoids simulating costly Markov chains as in classical computers. To achieve this, we first formulate the k-means objective as a quadratic energy function over binary variables and embed the clustering task into the quantum-physical system of D-Wave. By repeatedly measuring the quantum system, we sample high-probability solutions according to the Boltzmann distribution. Unlike previous approaches that only use the best solution and discard all other measurements, we utilize all samples to generate probabilistic solutions for the k-means problem, as shown in Figure 1. Due to temperature mismatch [68], AQCs sample from a modified posterior instead of the true posterior, which needs to be *re-calibrated* for the task at hand. We do so by estimating the posterior probability for each solution. This allows us to identify ambiguous points and provide alternative solutions. By leveraging QA, our approach respects the binary nature of the assignment problem.

We demonstrate our algorithm on a D-Wave quantum computer and also perform extensive experiments in simulation. On real data, we show that our approach can be utilized for distributions that do not strictly follow the initial assumptions and that it is suitable for identifying ambiguous images. Our primary contributions are:

- A quantum computing formulation of balanced k-means clustering that predicts calibrated confidence values and provides a set of alternative clustering solutions.
- A reparametrization approach used to calibrate posterior probabilities from samples that avoids exact tuning of the AQC sampling temperature.
- Extensive experiments on synthetic and real data showing the calibration of our approach both on simulations and on the D-Wave Advantage 2 QA prototype.

Developing better heuristics for the NP-hard challenges of clustering and UQ presents significant difficulties for classical approaches. However, we are optimistic that the ongoing progress in quantum computing technology will progressively enhance the effectiveness of our algorithms.

2. Related Work

Quantum computation. With the availability of quantum computers to the general research community [15, 28, 57, 58], the research interest in finding applications for such systems has considerably increased. In this context AQC [15, 47] provides a well-tangible starting point, even though many applications need a complete reformulation considering the architectural differences of a quantum computer. Current applications for AQC are optimization tasks in different fields [61, 62, 64, 66], including *quantum computer vision* [12], where a strong interest in finding quantum computing formulations has developed. These are often related to hard permutation problems [6, 7, 12] like tracking [77], graph, shape and point matching [6, 9, 60] or training binary (graph) neural networks [49]. In this context, Birdal et al. [12] evaluate the k-best solutions, however, without the energy-based formulation, only little improvement is achieved. Another tasks of interest for the community is model fitting to find camera parameters [34] or separating motion components [3], which can be formulated as a quantum-computing problem [3, 19, 20, 34, 37, 74] instead of consensus maximization.

Clustering. Clustering is a well-studied problem for quantum- as well as quantum-inspired algorithms [2]. It groups a dataset X into disjoint clusters $c_1, ..., c_K$, with each cluster containing points similar in the feature space. For example, k-means algorithm [39, 53] utilizes quadratic distances, favoring compactness, whereas dbscan [36] identifies local high-density regions. Quantum clustering methods use either the paradigm of AQC [4, 5, 50, 65] or circuit-based quantum computing such as Quantum clustering [43] that models clustering through the Schrödinger equation, where cluster centers are defined as the minima of its potential function. Casaña-Eslava et al. [18] extend this formulation with a probabilistic estimate of cluster memberships. These approaches use classical computation and circuit-based quantum computers that are fundamentally different from AQC and currently still on a smaller scale.

Closest to this work, Arthur and Date [4] present a balanced k-means clustering algorithm with predefined target size s_k suitable for an AQC and Nguyen et al. [65] use a similar formulation to cluster visual features. While our underlying algorithm follows a similar approach as [4, 65], they discard all but the best measurement. Our approach, however, utilizes all information by employing AQC as a sampler to generate probabilistic solutions of the clustering problem, rather than only using the best solution in an optimization framework. We particularly consider balanced clustering, as it forms the basis of several AQC algorithms [7, 12, 77] in computer vision.

Uncertainty quantification in clustering. Knowing the set of high-quality solutions and their confidence offers valuable insights for both low- and high-level tasks. E.g. at a low level, the probability of top solutions can help to determine the correct cluster count. Choosing a different number than the actual data process introduces ambiguity [17]: too many clusters cause overlap, while too few clusters spread points, reducing associated probability.

For high-level tasks, calibrated clustering solutions have the potential to enhance matching problems like multi-object tracking [21]. Following the AQC framework, tracking becomes a clustering problem with additional

constraints that represent the temporal relation between points [77], where each cluster corresponds to one track. The feature defining clusters can be visual similarities from re-identification [42, 67, 79] or spatial similarity [8]. The set of solutions models different trajectories due to occlusions or intersecting paths. In complex systems like autonomous vehicles, understanding multiple probable solutions helps to predict multimodal future trajectories [31, 46], to assess risks when taking actions. In a similar approach, AQC is used for feature matching [12] and uncertainty estimation potentially allows to discard ambiguous candidates during 3d reconstruction.

3. Adiabatic Quantum Computing

Quantum computing is a fundamentally new approach, that utilizes the state of a quantum system to perform computations. In contrast to a classical computer, the state is probabilistic and described by its wave function, which enables the use fundamental properties of quantum systems like superposition and entanglement. Quantum algorithms that can efficiently run on these systems allow to solve previously infeasible problems, including the well-known prime factorization by Shor's algorithm [71]. In the following, the most important basics of quantum computing are explained.

Qubit. Qubits are the building block of a quantum computer. In contrast to the classical bit, which is in either of its two basis states $|0\rangle, |1\rangle$, the qubit can exist in a state of superposition, which is a linear combination of any two basis states $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$. The two complex-valued factors α, β describing the superposition are called amplitudes.

Entanglement. To perform meaningful operations, a quantum computer operates on a system of multiple qubits. If the qubits are independent of each other, the joint state of the system can be computed as the tensor-product of all involved qubit states. If the qubits of a system are entangled [35, 70], it is is not possible to decompose the joint state into the tensor-product of each separate qubit state. Therefore, measuring the state of one qubit in a entangled system, affects the state of the other qubits. This is a fundamental property of quantum mechanics and plays a crucial role for the capabilities of quantum computing.

Measurement. During computation on the quantum computer, the state of the system can be any valid superposition of basis states. Nevertheless, a measurement of the system always results in a single state of the measurement basis, which is referred to as wave-function collapse [73]. The probability of measuring a state is the respective squared amplitude. In the single qubit case, this corresponds to

$$p(|0\rangle) = |\alpha|^2 \qquad p(|1\rangle) = |\beta|^2. \tag{1}$$

In contrast to all other operations in quantum computing, this is not reversible.

Adiabatic Quantum Computing. AQC, which is used in this work, is a quantum computing paradigm where the state of a quantum system is modified by performing an adiabatic transition. Current hardware implementations such as the D-wave systems [15] follow this approach by implementing QA to solve quadratic unconstrained binary optimization (QUBO). They are based on the Ising model [45, 47], which describes the configuration of a set of interacting particles that all carry an atomic spin s_i.

The spin can either be +1 or -1 and the particles are coupled by interactions J_{ij} as well as influenced individually by a transversal magnetic field h_i. The energy of this system is described by its Hamiltonian function

$$H(s) = -\sum_i \sum_j J_{ij} s_i s_j - \sum_i h_i s_i. \tag{2}$$

Thus, finding the lowest energy state or ground state of the Ising model corresponds to solving the QUBO defined by the Hamiltonian function. This relation is used in the QA, where the system of qubits implements an Ising model H_T that represents the QUBO of interest.

The lowest energy state is found by following the adiabatic theorem [14]. Starting with an initial system in its ground state described by a Hamiltonian H_0, the adiabatic theorem states that during a sufficiently slow change of the Hamiltonian the system never leaves its ground state. The change of the Hamiltonian is called an adiabatic transition and is often performed with a linear schedule

$$H(t) = H_0(1 - \frac{t}{T_a}) + H_T \frac{t}{T_a} \tag{3}$$

over the annealing-time T_a and allows to solve an optimization problem with the Hamiltonian H_T.

While in an ideal noise-free case, the system stays in its ground state, any real system is embedded in a temperature bath that can induce a change to a higher energy state. The distribution of measured final states will then follow the Boltzmann distribution with temperature T

$$p(s) = \exp[-H(s)/T] / \sum_{s'} \exp[-H(s')/T], \tag{4}$$

where $p(s)$ describes the probability of finding the system in state s and s' are all possible states. In this work, we use this property, to sample from the Boltzmann distribution corresponding to the energy-based model of clustering.

4. Balanced Quantum K-Means

In the following, we define an *energy-based model* (EBM) that directly connects to AQC. Based on this, we derive the energy function for k-means and demonstrate how an AQC can sample from the corresponding distribution.

EBM Inference. We use the variable $Z \in \mathbb{Z}_2^N$, to represent our binary parameters. The posterior density of interest follows from the Bolzmann distribution as:

$$\pi(Z) := p(Z|X) \propto \exp(-E(Z,X)), \tag{5}$$

where X denotes the observations, a.k.a. the collection of all specified data and E is called the *potential energy*:

$$E(Z,X) := -(\log p(X|Z) + \log p(Z)). \tag{6}$$

Following [11, 12], we consider the probabilistic inference, where we will be interested in the following quantities:
I Maximum a-posteriori (MAP):

$$\hat{Z} = \arg\max_{Z \in \mathbb{Z}_2^N} \log p(Z|X) \tag{7}$$

where $\hat{Z}$ denotes the entirety of the sought parameters.
II The full posterior distribution: $p(Z|X) \propto p(X,Z)$.
Both of these problems are very challenging and cannot be directly addressed by standard methods such as gradient descent (problem I) or standard MCMC methods (problem II). The difficulty in these problems is mainly originated by the fact that the posterior density is non-log-concave (i.e. the negative log-posterior is non-convex) and any algorithm that aims at solving one of these problems should be able to operate in the particular space of binary variables. Nevertheless, The MAP estimate is often easier to obtain via a Quantum annealer and useful in practice [7, 12]. On the other hand, samples from the full posterior can provide important additional information, such as *uncertainty*. Not surprisingly, the latter is a much harder task, especially considering the discrete nature of our problem. Unfortunately, available QAs such as D-Wave [58] are not directly capable of tackling the second problem since they sample from a modified posterior and not the true one. Our work fills this gap for the task of balanced K-means clustering using quantum annealing as a direct way to sample from a physical system that follows the Boltzmann distribution [68] and recomputes the posterior from the found solutions to ensure callibration.

Clustering as QUBO. We now focus on obtaining a single point-estimate as a solution to problem I. To solve the clustering problem using AQC, a QUBO formulation of the K-means energy is required. We use a variation of the one-hot encoding approach [27] to formulate the QUBO of k-means, with cost terms similar to Arthur and Date [4]. It uses a matrix $Z \in \{0,1\}^{K \times I}$ to encode the cluster assignment of I samples to K clusters. Each row corresponds to one of K clusters and each column to one of I samples. An entry $Z_{ki} = 1$ indicates that the sample x_i belongs to cluster c_k. As each sample needs to be assigned to a single cluster, the sum of each column of Z needs to satisfy the constraint $\sum_k Z_{ki} = 1 \; \forall i$. The implementation of constrained clustering, where each cluster has a fixed size s_k furthermore requires the row constraints $\sum_i Z_{ki} = s_k \; \forall k$ on Z. The total energy of a solution $E(Z,X)$ can be separated into terms $e(i,j,k)$ modeling the energy of assigning the pair of samples x_i and x_j to the same cluster c_k.

$$E(Z,X) = \sum_k \sum_i \sum_j Z_{ki} Z_{kj} e(i,j,k) + E(Z) \tag{8}$$

where $E(Z)$ models the row and column constraints as an indicator function. By vectorizing Z in row-major order as $z = \text{vec}(Z)$, the energy can be rewritten in matrix form as

$$E(Z,X) = E(X|Z) + E(Z) = z^\intercal \, Q \, z + E(Z), \tag{9}$$

where Q is a block diagonal matrix with blocks $Q_0, ..., Q_K$ and $Gz = d$ corresponds to the matrix formulation of the constraints. Each block Q_k of Q is a square matrix that contains the energy $e(i,j,k)$ at $Q_{k,ij}$.

To be solved on the AQC, the potential energy needs to be formulated in a QUBO, which cannot exactly implement $E(Z)$ to model the constraints and is circumvented by using Lagrangian multipliers, leading to the MAP estimate

$$\hat{z} = \arg\min_z z^\intercal Q' z + b'^\intercal z \tag{10}$$

with constraints $Q' = Q + \lambda G^\intercal G$ and $b' = -2\lambda G^\intercal b$. To avoid a mixed discrete and continuous optimization problem, a quadratic penalty reformulation $\lambda ||Gz - d||_2^2$ is chosen. In contrast to a linear penalty approach, where the multiplier λ needs to be optimized, our selection only requires a sufficiently high λ, as all constraint violations result in an increased penalty term. With such selection, the penalty term evaluates to zero if all constraints are fulfilled, and thus, the minimizer $\hat{z}$ of the modified optimization problem is a minimizer of the original optimization problem. Besides modeling balanced clustering, further constraints can be introduced using Lagrangian multipliers. This allows to represent a wide range of tasks as QUBO clustering problems solvable with AQC in the same way.

4.1. Probabilistic Quantum Clustering

The MAP estimate was obtained as the lowest energy solution measured during sampling. We now consider approximating the posterior distribution to quantify the confidence of these solutions matching the actual ground truth. Instead of relying on a sequential Markov Chain Monte Carlo (MCMC) method as done in a plethora of classical algorithms [29], we will approximate the posterior distribution by recomputing the Boltzmann distribution from the set of likely AQC solutions. As the likely solutions are already sampled during quantum annealing, given that the objective follows the EBM, the posterior distribution can be estimated without any large additional computational overhead.

Note, this is unlike MCMC, which incurs significant computational load.

Our clustering formulation employs a mixture of Gaussians to explain the observations and to define the energy function $E(X|Z)$. Each sample in a cluster c_k is modeled as a sample drawn from a Gaussian distribution $\mathcal{N}(\mu_k, I)$ with mean μ_k and identity covariance similar to k-means. As shown in Section 4, we are interested in the posterior distribution over possible assignments Z:

$$p(Z|X) = \frac{p(X|Z)p(Z)}{\sum_{Z'} p(X|Z')} = \frac{p(X|Z)p(Z)}{A}. \tag{11}$$

While this approach provides a probabilistic estimate by jointly modeling information about the possible cluster configurations and the distribution of data points, it is often intractable to evaluate due to the partition function $\sum_{Z'} p(X|Z')$, the sum over all possible solutions Z', which incurs exponential cost in the number of samples. To overcome this, we utilize an AQC that samples directly from the corresponding Boltzmann distribution, which we parameterize according to the probabilistic clustering problem.

Data Model. Determining the potential energy and thus, cost function of the QUBO is a design choice of the algorithm. For our probabilistic approach, a well-defined data distribution, that forms the basis of the cost function is required. While many tasks approached in quantum computer vision, such as tracking [77] or synchronization [6, 12], costs are based on learned metrics or heuristics, they can also be trained to reflect the properties required in our approach.

We therefore, follow the mixture of Gaussian model, where each cluster generates samples from a normal distribution. With the independence of observations and clusters, given the distribution parameters, the likelihood of the joint observations for an assignment Z is given as the product of the individual likelihoods

$$f(X|Z) = \prod_{k=1}^{K} f(X|Z_k) = \prod_{k=1}^{K} \prod_{i \in Z_k} f(x_i|Z_k) \tag{12}$$

where the likelihood $f(x_i|Z_k)$ corresponds to a Gaussian distribution that follows $\mathcal{N}(\mu_k, I)$ and d is the dimensionality of the space. This result can be used to formulate the energy-based model with the energy function

$$E(X|Z) = \sum_{k=1}^{K} \sum_{i \in Z_k} \frac{1}{2}(x_i - \mu_k)^\intercal I (x_i - \mu_k). \tag{13}$$

Using the energy function to rewrite the posterior distribution leads to the Boltzmann distribution from Equation 4

$$p(Z|X) = \frac{\exp\left[-(E(X|Z) + E(Z))\right]}{\sum_{Z'} \exp\left[-E(X|Z')\right]}. \tag{14}$$

In this formulation, the energy of the assignment $E(Z)$ corresponds to the prior $p(Z)$, which models feasible and infeasible solutions by an indicator function. In the case of balanced clustering, this corresponds to allowing assignments that have each point assigned to exactly one cluster and a cluster size according to s_k. When the AQC is used as an optimizer, finding the the lowest energy solution corresponds to the MAP as shown in Equation 7

As the AQC qubit system is an Ising model, it requires formulating $E(X|Z)$ and $E(Z)$ quadratic in the optimization variables Z, enabling the joint discovery of assignments and cluster means. This is achieved by using the maximum likelihood estimator of the mean $\mu_k = \frac{1}{s_k} \sum_i Z_{ki} x_i$, resulting in a quadratic energy formulation that fits the Ising model in Equation 10

$$E_k(X|Z) = \frac{1}{s_k} \sum_i \sum_j Z_{ki} Z_{kj} \, (x_i - x_j)^\intercal (x_i - x_j). \tag{15}$$

The second energy term $E(Z)$ in Equation 14, modeling the prior distribution over possible assignments cannot exactly be embedded in the Ising model and is approximated using Lagrange multipliers. Importantly, this does not influence the energy of feasible solutions relevant in the posterior distribution, as the penalty evaluates to zero for these.

Posterior recomputation. While ideally the measurements on AQC are direct samples from the Boltzmann distribution [25, 33], it requires solving a range of challenges that prohibit a direct use in our scenario [68]: Mapping between the cost function and the physical system implemented on the AQC requires Hamiltonian scaling [68]. Estimating the scaling factor is nontrivial [63] and prohibitive for using samples directly in many cases. Additionally, hardware limitations including imperfection of the processor and the spin-bath polarization effect [68] prevent the AQC to sample the Boltzmann distribution exactly. Finally, as the energy term $E(Z)$ can only be implemented using the penalty method, the sampling density is influenced by solutions not fulfilling the constraints. We compensate for these limitations by evaluating the energy of all measured feasible solutions Z' and recompute $p(Z|X)$ by evaluating the analytic partitioning function over these, which only requires sampling solutions at a sufficiently high temperature.

Coresets. Having the set of the most likely clustering solutions Z available together with their posterior probability $P(Z|X)$ allows to find an assignment Z^* of a subset of points that solves the clustering problem with increased probability $P(Z^*|X)$. Such a set can be found by using Algorithm 1, which implements a greedy approach that disregards points that disagree between different clustering solutions. With a sufficiently well-sampled Boltzmann distribution, the resulting coreset assignment Z^* with the corresponding probabilistic estimate $P(Z^*|X)$ is still calibrated.

Algorithm 1 CoresetSearch

$Z^* \leftarrow Z_0$
$p \leftarrow P(Z_0|X)$
$i \leftarrow 1$
while $p \leq p_{\text{min}}$ **do**
  $Z_i' \leftarrow align(Z_i, Z^*)$
  ▷ Find the cluster permutation that minimizes the number of points assigned to different clusters in Z_i and Z^*.
  $Z^* \leftarrow Z^* \cap Z_i'$
  ▷ Remove points assigned to different clusters.
  $p \leftarrow p + P(Z_i|X)$
  $i \leftarrow i + 1$
end while
return Z^*

Lagrange multiplier optimization. Using the quadratic penalty method to implement the constraints requires finding suitable Lagrangian multipliers λ. Even though a very high multiplier theoretically guarantees finding a feasible solution, it also deteriorates the conditioning of the optimization problem. Therefore, a suitable Lagrangian multiplier lifts the cost of any constraint violation above all relevant solutions of the clustering problem, while keeping them low enough to avoid scaling the total energy of the problem up considerably. To estimate the multipliers, we follow an iterative procedure as also proposed in [77].

In an initial step, balanced classical k-means [56] is used to find a feasible clustering solution. This solution is used to estimate Lagrangian multipliers that avoid any first-order violation of the constraints. In subsequent iterative optimization steps, the problem is solved in simulation, to find multipliers that result in a well-conditioned problem.

Such optimization procedure is crucial due to the low fidelity of the current generation of QAs, which requires careful engineering of the problem energy. Therefore, we expect this procedure to become of reduced importance with the progress of quantum computing.

5. Experiments and Results

We perform experiments on synthetic as well as real data to verify the efficacy of our method in finding the set of high-probability solutions and in estimating calibrated confidence scores. The experimental scenarios are solved with QA, Simulated Annealing (SIM), and exact exhaustive search using the presented energy formulation and with k-means as a baseline method. This further allows us to understand the limitations and required work when deploying the approach to real quantum computers.

Implementation details. **Quantum annealing** (QA) experiments are performed on the D-Wave Advantage 2 Prototype 1.1 [58]. The system offers 563 working qubits, each

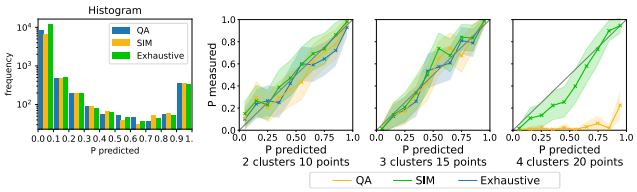


Figure 2. Evaluation of the calibration and distribution for QA, SIM and exhaustive search in tasks with 2 to 4 clusters, each with 5 points. All results are generated with 1,000 problems in each scenario and 5,000 measurements for each clustering problem.

connected with up to 20 neighbors. For each clustering problem 5000 measurements are performed, each with 50μs annealing time. Due to the strong compute-time limits on an QA, all Lagrangian optimization steps are performed with SIM, before measuring the final results on the QA.

Simulated annealing (SIM) provided by D-Wave is used for larger scale comparisons. Similar to QA, we perform 5000 runs for each clustering problem. We reduce the number of sweeps performed in each run to 30, which allows us to sample the Boltzmann distribution at a sufficiently high temperature in most scenarios, making it comparable to QA.

Exact exhaustive search is used as a reference method to validate the energy-based formulation on small problems. By iterating all feasible solutions, the lowest energy solution is guaranteed to be found and the partitioning function is computed exactly.

K-Means clustering with a balanced cluster constraint [56] forms the baseline for our approach. We run the algorithm until convergence for a maximum of 1000 iterations. While this solution does not provide a probabilistic estimate, it is useful to assess the relative clustering performance.

Data for the quantitative evaluation of our method is synthetically generated. For each clustering problem a total of I points are sampled from a separate normal distribution for each of K clusters. The centroids are randomly drawn, such that the distance between each pair of clusters lies within a predefined range $[d_{\text{min}}, d_{\text{max}}]$. For each experiment a total of L clustering tasks is generated. This allows us to evaluate the calibration metrics over a large value range. For all experiments that directly compare methods, identical clustering tasks are used.

Further results are provided for the IRIS dataset [38] and a dataset of images containing ambiguous objects that are to be identified. The IRIS contains 50 samples of 4 features in 3 classes. We randomly subsample the points and dimensions to generate the parameters required for our experiments.

Clustering metrics are computed using the available ground-truth clusters. We evaluate 4 standard metrics: The accuracy, which measures the ratio of clustering solutions that are identical to the ground-truth. Completeness [69] measures the ratio of points from a single cluster being

Solver Method	Accuracy	Completeness	Adjusted Rand Index	Fowlkes-Mallows Score	Accuracy	Completeness	Adjusted Rand Index	Fowlkes-Mallows Score	Accuracy	Completeness	Adjusted Rand Index	Fowlkes-Mallows Score
	15 Points, 3 Clusters, 2 Dim				30 Points, 3 Clusters, 2 Dim				45 Points, 3 Clusters, 2 Dim			
SIM	**56.4**±1.6	**79.5**±0.8	**74.7**±1.0	**81.9**±0.7	**38.6**±1.5	**74.2**±0.8	**73.3**±0.9	**81.6**±0.6	50.8±1.6	**86.2**±0.5	**87.3**±0.5	**91.4**±0.3
K-means	51.3±1.6	75.0±0.9	68.8±1.1	77.7±0.8	37.0±1.5	71.9±0.8	70.2±0.9	79.4±0.6	**52.8**±1.1	85.9±0.4	86.6±0.4	90.9±0.3
	15 Points, 3 Clusters, 2 Dim				10 Points, 2 Clusters, 2 Dim				20 Points, 4 Clusters, 4 Dim			
QA	56.1±1.6	79.4±0.8	74.6±1.0	**81.9**±0.7	**74.3**±1.4	**80.2**±1.1	**79.6**±1.1	**88.7**±0.6	12.4±1.0	51.2±0.8	34.0±1.0	47.9±0.8
SIM	**56.4**±1.6	**79.5**±0.8	**74.7**±1.0	**81.9**±0.7	74.1±1.4	80.0±1.1	79.5±1.1	88.6±0.6	**32.4**±1.5	**70.4**±0.8	**59.2**±1.1	**67.8**±0.8
K-means	51.3±1.6	75.0±0.9	68.8±1.1	77.7±0.8	70.1±1.4	76.3±1.2	75.4±1.2	86.3±0.7	20.9±1.3	61.9±0.8	47.4±1.0	58.5±0.8
Exhaustive	**56.4**±1.6	**79.5**±0.8	**74.7**±1.0	**81.9**±0.7	**74.3**±1.4	**80.2**±1.1	**79.6**±1.1	**88.7**±0.6	-	-	-	-

Table 1. Synthetic data results for our approach (QA, SIM, exhaustive) and k-means. All numbers in % with standard error of the mean.

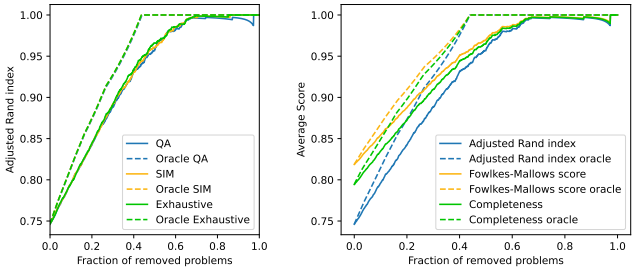
grouped together. The adjusted Rand score [44] compares all pairs of points in the ground truth and prediction and the Fowlkes-Mallows index [40] combines precision and recall into a single score.

5.1. Results

Calibration performance. We evaluate the calibration of our method on a synthetic clustering scenario with 3 clusters and 15 points using QA, SIM and exhaustive search on 1000 tasks. First all clustering solutions Z are accumulated in bins according to their estimated posterior probability $P(Z|X)$. This process also includes all sampled non-optimal but feasible solutions. Figure 2 shows the resulting histogram of solutions and the ratio of correct solutions in each bin after accumulation. It thus evaluates calibration, where the ideal case has mean predicted probabilities on the diagonal. We find our approach to generate well-calibrated probabilities in both simulation and when using QA for the two smaller experiments for up to 15 points. Due to the problem QUBO size, only SIM is able to find the correct solutions on the largest example.

Clustering performance. We evaluate the performance of our clustering formulation using synthetic data in Table 1. The upper rows show results for 3 clusters with an increasing number of points and 1000 tasks/setting for SIM and k-means. Due to the problem size, it cannot be solved using QA and exhaustive search. Our formulation with SIM outperforms k-means on the small tasks, however, the difference vanishes with increasing problem size. This can be attributed to the globally optimal solution our approach is optimizing for. Ideally, the global solution is at least as good as the k-means solution, however, an increasing problem size also increases the complexity of the QUBO, which explains the dropping performance on large problems.

For the largest clustering problem with 3 clusters and 45 points, k-means provides the best accuracy with more correct solutions compared to SIM. However, the clustering quality metrics are higher for SIM. This indicates that our formulation is able to find a better solution in cases where the problem is not solved correctly by SIM and k-means.



(a) Solver comparison for the adjusted Rand index.

(b) Metric comparison for the QA solver.

Figure 3. Sparsification plots of clustering metrics.

The lower rows in Table 1 show settings with 2,3, and 4 clusters, each containing 5 points, again with 1000 tasks/setting. For the two smaller scenarios QA on the D-wave Advantage 2 provides results close or identical to SIM and exhaustive search. For the largest scenario with 20 points in 4 clusters, QA loses performance.

The quality of predicted posterior probabilities can be further evaluated by linking them to the clustering metrics and sparsifying the set of tasks. Starting with metrics over the whole set of clustering tasks, we remove tasks according to increasing probability and evaluate the metric over the remaining set. For an ideal predictor, this removes the lowest-performing tasks first. In Figure 3a this is done for the adjusted Rand index with *different solvers* and in Figure 3b QA with *different metrics*. The solid lines show the sparsification plots using the predicted probabilities and the dashed lines show the same plots for an oracle method that generates the best possible ordering based on the metrics.

In Figure 3a it becomes apparent that QA, SIM and exhaustive search perform close to each other over most of the value range. Nevertheless, for a high sparsification with more than 80% of the tasks removed, QA shows a drop in performance compared to the other methods. This is caused by tasks where only a single, but incorrect solution is found, which gets assigned a posterior probability of $P(Z|X) = 1.0$. In such cases, the Boltzmann distribution has not been sampled sufficiently well, either because of too few measurements or because of a low effective sam-

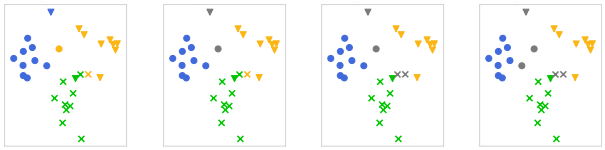
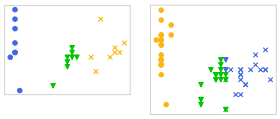


Figure 4. Visualization of coresets for synthetic data with the probability for each of the determined pointsets.



(a) QA. (b) SIM.

Figure 5. Qualitative results on the IRIS dataset.

	Accuracy	Completeness	Adjusted Rand Index	Fowlkes-Mallows Score
QA	47.2	81.7	76.8	83.4
SIM	47.1	81.7	76.7	83.4
K-means	47.0	80.6	75.5	82.5
Exhaustive	47.1	81.8	76.8	83.5

Table 2. Performance on IRIS subsets (3 clusters/15 points).

pling temperature. As SIM does not show this behavior the source can likely be traced back to current limitation of the quantum computer.

Coresets. Qualitative examples for the coresets generated with Algorithm 1 are depicted in Figure 4. The shape of each point represents the ground truth class and the color the assigned cluster. Starting with the most likely solution having a predicted probability of $p(Z|X) = 0.61$ on the left, each plot shows one additional step of the algorithm, which successively removes points from the solution, indicated by plotting them in Grey. The illustration demonstrates that the CoresetSearch algorithm is able to generate well-separated clusters from the probabilistic predictions.

Real-World Datasets While our method assumes an identity covariance in the data, it can be applied to other distributions, which we evaluate on the widely used IRIS dataset as well as a set of self-collected images. Clustering metrics for experiments on IRIS using 3 clusters with 5 points each are provided in Table 2 and show that our formulation using QA and SIM is competitive with k-means. Figures 5a and 5b show differently sized qualitative examples from the dataset solved using our formulation with QA and SIM respectively. On the full IRIS dataset, SIM and k-means achieve identical results with a Completeness of 77.7%, Adjusted Rand index of 78.6% and further results provided in the supplementary. The results from Algorithm 1 using results from SIM are provided in Figures 6 and 7 for IRIS and our dataset respectively. They show that even for a distribution mismatch the coresets can provide meaningful results for removing ambiguous samples.

To demonstrate that meaningful ambiguous samples can be identified in a high-dimensional space present in image data, we collected publicly available images of cars, boats and cars towing boats. Image features were extracted using VGG [72] and subsequently clustered using our approach. Similar to IRIS, the distribution does not match our assumptions, however, we are still able to identify ambiguous images that contain cars towing a boat by using Algorithm 1. The first coreset is depicted in Figure 7 where differently colored frames indicate the cluster assignment and Grey frames mark ambiguous samples. Clustering is performed on the 1000-dimensional feature and images are plotted at their 2D projection generated using UMAP [59]. Our approach correctly identifies the images that contain a car towing a boat and thus, cannot be uniquely assigned to either cluster.

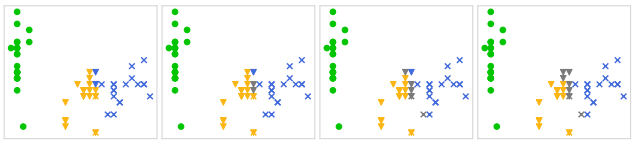


Figure 6. Coresets generated using SIM on the IRIS dataset with 60 points.

Figure 7. First coreset using SIM on our collected dataset.

6. Conclusion

In this work, we proposed a probabilistic clustering approach based on sampling k-means solutions using AQC. By using all valid measurements, calibrated confidence scores are computed at little cost and solutions are competitive to an iterative balanced k-means approach. We evaluated our method on synthetic as well as real data using simulation, exhaustive search as well as the D-Wave Advantage 2 prototype QA to explore the potential of quantum computing in machine learning and computer vision.

While the approach is still limited in the problem size, quantum computing enables a fundamentally different approach to clustering that can provide additional information that is costly to compute otherwise. Nevertheless, even with the current progress and potential to scale to real-world problems, more work is required to adapt existing problem formulations, such that the full capability of quantum computing can be effectively used.

Acknowledgement: This work was funded by Toyota Motor Europe via the research project TRACE Zürich.

References

[1] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021. 1

[2] Esma Aïmeur, Gilles Brassard, and Sébastien Gambs. Quantum clustering algorithms. In *Proceedings of the 24th International Conference on Machine Learning - ICML '07*, pages 1–8, Corvalis, Oregon, 2007. ACM Press. 2

[3] Federica Arrigoni, Willi Menapace, Marcel Seelbach Benkner, Elisa Ricci, and Vladislav Golyanik. Quantum Motion Segmentation. In *Computer Vision – ECCV 2022*, pages 506–523. Springer Nature Switzerland, Cham, 2022. 2

[4] Davis Arthur and Prasanna Date. Balanced k-means clustering on an adiabatic quantum computer. *Quantum Information Processing*, 20(9):294, 2021. 2, 4

[5] Christian Bauckhage, E. Brito, K. Cvejoski, C. Ojeda, Rafet Sifa, and S. Wrobel. Ising Models for Binary Clustering via Adiabatic Quantum Computing. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 3–17, Cham, 2018. Springer International Publishing. 2

[6] M. Benkner, Z. Lahner, V. Golyanik, C. Wunderlich, C. Theobalt, and M. Moeller. Q-match: Iterative shape matching via quantum annealing. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7566–7576, Los Alamitos, CA, USA, 2021. IEEE Computer Society. 2, 5

[7] Marcel Seelbach Benkner, Vladislav Golyanik, Christian Theobalt, and Michael Moeller. Adiabatic Quantum Graph Matching with Permutation Matrix Constraints. In *2020 International Conference on 3D Vision (3DV)*, pages 583–592, Fukuoka, Japan, 2020. IEEE. 2, 4

[8] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple Online and Realtime Tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468, 2016. 3

[9] Harshil Bhatia, Edith Tretschk, Zorah Lähner, Marcel Seelbach Benkner, Michael Moeller, Christian Theobalt, and Vladislav Golyanik. CCuantuMM: Cycle-Consistent Quantum-Hybrid Matching of Multiple Shapes. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1296–1305, Vancouver, BC, Canada, 2023. IEEE. 2

[10] Jacob D. Biamonte and Peter J. Love. Realizable Hamiltonians for Universal Adiabatic Quantum Computers. *Physical Review A*, 78(1):012352, 2008. 2

[11] Tolga Birdal and Umut Simsekli. Probabilistic permutation synchronization using the riemannian structure of the birkhoff polytope. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11105–11116, 2019. 4

[12] Tolga Birdal, Vladislav Golyanik, Christian Theobalt, and Leonidas Guibas. Quantum Permutation Synchronization. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13117–13128, Nashville, TN, USA, 2021. IEEE. 2, 3, 4, 5

[13] Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*. Springer, 2006. 1

[14] M. Born and V. Fock. Beweis des Adiabatensatzes. *Zeitschrift für Physik*, 51(3):165–180, 1928. 3

[15] P. I. Bunyk, E. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, and J. Whittaker. Architectural considerations in the design of a superconducting quantum annealing processor. *IEEE Transactions on Applied Superconductivity*, 24(4):1–10, 2014. 2, 3

[16] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep Clustering for Unsupervised Learning of Visual Features. In *Computer Vision – ECCV 2018*, pages 139–156. Springer International Publishing, Cham, 2018. 1

[17] Raúl V. Casaña-Eslava, Paulo J. G. Lisboa, Sandra Ortega-Martorell, Ian H. Jarman, and José D. Martín-Guerrero. A Probabilistic framework for Quantum Clustering, 2019. 2

[18] Raúl V. Casaña-Eslava, Paulo J. G. Lisboa, Sandra Ortega-Martorell, Ian H. Jarman, and José D. Martín-Guerrero. Probabilistic quantum clustering. *Knowledge-Based Systems*, 194:105567, 2020. 2

[19] Yulin Chi, Jieshan Huang, Zhanchuan Zhang, Jun Mao, Zinan Zhou, Xiaojiong Chen, Chonghao Zhai, Jueming Bao, Tianxiang Dai, Huihong Yuan, Ming Zhang, Daoxin Dai, Bo Tang, Yan Yang, Zhihua Li, Yunhong Ding, Leif K. Oxenløwe, Mark G. Thompson, Jeremy L. O'Brien, Yan Li, Qihuang Gong, and Jianwei Wang. A programmable qudit-based quantum processor. *Nature Communications*, 13(1): 1166, 2022. 2

[20] Tat-Jun Chin, David Suter, Shin-Fang Ch'ng, and James Quach. Quantum Robust Fitting. In *Computer Vision – ACCV 2020*, pages 485–499. Springer International Publishing, Cham, 2021. 2

[21] Hsu-kuang Chiu, Jie Li, Rares Ambrus, and Jeannette Bohg. Probabilistic 3D multi-modal, multi-object tracking for autonomous driving. *IEEE International Conference on Robotics and Automation (ICRA)*, 2021. 2

[22] Adam Coates and Andrew Y. Ng. Learning Feature Representations with K-Means. In *Neural Networks: Tricks of the Trade*, pages 561–580. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. 1

[23] G.B. Coleman and H.C. Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979. 1

[24] Gregory F Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial intelligence*, 42(2-3):393–405, 1990. 1

[25] D-Wave Systems Inc. Performance advantage in quantum Boltzmann sampling. Technical report, D-Wave Systems Inc., 2017. 5

[26] Sanjoy Dasgupta. The hardness of k-means clustering. 2008. 1

[27] Prasanna Date, Davis Arthur, and Lauren Pusey-Nazzaro. QUBO formulations for training machine learning models. *Scientific Reports*, 11(1):10029, 2021. 4

[28] S. Debnath, N. M. Linke, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe. Demonstration of a small pro-

grammable quantum computer with atomic qubits. *Nature*, 536(7614):63–66, 2016. 2

[29] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436, 2006. 1, 4

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977. 1

[31] Nachiket Deo and Mohan M. Trivedi. Multi-Modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs. In *IEEE Intelligent Vehicles Symposium (IV)*, 2018. 3

[32] Nameirakpam Dhanachandra, Khumanthem Manglem, and Yambem Jina Chanu. Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, 54:764–771, 2015. 1

[33] Vivek Dixit, Raja Selvarajan, Muhammad A. Alam, Travis S. Humble, and Sabre Kais. Training Restricted Boltzmann Machines With a D-Wave Quantum Annealer. *Frontiers in Physics*, 9, 2021. 1, 5

[34] Anh-Dzung Doan, Michele Sasdelli, David Suter, and Tat-Jun Chin. A Hybrid Quantum-Classical Algorithm for Robust Fitting. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 417–427, New Orleans, LA, USA, 2022. IEEE. 2

[35] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, 47(10):777–780, 1935. 3

[36] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, Oregon, 1996. AAAI Press. 2

[37] Matteo Farina, Luca Magri, Willi Menapace, Elisa Ricci, Vladislav Golyanik, and Federica Arrigoni. Quantum Multi-Model Fitting. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13640–13649, Vancouver, BC, Canada, 2023. IEEE. 2

[38] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2):179–188, 1936. 6

[39] E. Forgy. Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification. *Biometrics*, 21(3): 768–769, 1965. 2

[40] E. B. Fowlkes and C. L. Mallows. A Method for Comparing Two Hierarchical Clusterings. *Journal of the American Statistical Association*, 78(383):553–569, 1983. 7

[41] Wenchong He and Zhe Jiang. A survey on uncertainty quantification methods for deep neural networks: An uncertainty source perspective. *arXiv preprint arXiv:2302.13425*, 2023. 1

[42] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof. Person Re-identification by Descriptive and Discriminative Classification. In *Image Analysis*, pages 91–102. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. 3

[43] David Horn and Assaf Gottlieb. The method of quantum clustering. In *Advances in Neural Information Processing Systems*. MIT Press, 2001. 2

[44] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985. 7

[45] Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925. 3

[46] Chiyu "Max" Jiang, Andre Cornman, Cheolho Park, Benjamin Sapp, Yin Zhou, and Dragomir Anguelov. MotionDiffuser: Controllable Multi-Agent Motion Prediction Using Diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9644–9653, 2023. 3

[47] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse Ising model. *Physical Review E*, 58 (5):5355–5363, 1998. 1, 2, 3

[48] Margret Keuper, Siyu Tang, Bjoern Andres, Thomas Brox, and Bernt Schiele. Motion Segmentation & Multiple Object Tracking by Correlation Co-Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):140–153, 2020. 1

[49] Maximilian Krahn, Michelle Sasdelli, Fengyi Yang, Vladislav Golyanik, Juho Kannala, Tat-Jun Chin, and Tolga Birdal. Projected stochastic gradient descent with quantum annealed binary gradients. *arXiv preprint arXiv:2310.15128*, 2023. 2

[50] Vaibhaw Kumar, Gideon Bass, Casey Tomlin, and Joseph Dulny III. Quantum Annealing for Combinatorial Clustering. *Quantum Information Processing*, 17(2):39, 2018. 2

[51] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pages 2169–2178, 2006. 1

[52] Stuart Lloyd. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2):129–137, 1982. 1

[53] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. 2

[54] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. Oakland, CA, USA, 1967. 1

[55] Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. The planar k-means problem is NP-hard. *Theoretical Computer Science*, 442:13–21, 2012. 1

[56] Mikko I. Malinen and Pasi Fränti. Balanced K-Means for Clustering. In *Structural, Syntactic, and Statistical Pattern Recognition*, pages 32–41. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. 1, 6

[57] Catherine McGeoch and Pau Farré. The Advantage System: Performance Update. Technical report, D-Wave Systems Inc., 2021. 2

[58] Catherine McGeoch, Pau Farre, and Kelly Boothby. The D-Wave Advantage2 Prototype. *Technical Report*, 2022. 2, 4, 6

[59] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *ArXiv e-prints*, 2018. 8

[60] Natacha Kuete Meli, Florian Mannel, and Jan Lellmann. An Iterative Quantum Approach for Transformation Estimation from Point Sets. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 519–527, 2022. 2

[61] Samuel Mugel, Mario Abad, Miguel Bermejo, Javier Sánchez, Enrique Lizaso, and Román Orús. Hybrid quantum investment optimization with minimal holding period. *Scientific Reports*, 11(1):19587, 2021. 2

[62] Vikram Khipple Mulligan, Hans Melo, Haley Irene Merritt, Stewart Slocum, Brian D. Weitzner, Andrew M. Watkins, P. Douglas Renfrew, Craig Pelissier, Paramjit S. Arora, and Richard Bonneau. Designing Peptides on a Quantum Computer, 2020. 2

[63] Jon Nelson, Marc Vuffray, Andrey Y. Lokhov, Tameem Albash, and Carleton Coffrin. High-Quality Thermal Gibbs Sampling with Quantum Annealing Hardware. *Physical Review Applied*, 17(4):044046, 2022. 5

[64] Florian Neukart, Gabriele Compostella, Christian Seidel, David von Dollen, Sheir Yarkoni, and Bob Parney. Traffic Flow Optimization Using a Quantum Annealer. *Frontiers in ICT*, 4:29, 2017. 2

[65] Xuan Bac Nguyen, Benjamin Thompson, Hugh Churchill, Khoa Luu, and Samee U. Khan. Quantum Vision Clustering, 2023. 1, 2

[66] Masayuki Ohzeki, Akira Miki, Masamichi J. Miyama, and Masayoshi Terabe. Control of Automated Guided Vehicles Without Collision by Quantum Annealer and Digital Devices. *Frontiers in Computer Science*, 1:9, 2019. 2

[67] Sakrapee Paisitkriangkrai, Chunhua Shen, and Anton Van Den Hengel. Learning to rank in person re-identification with metric ensembles. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1846–1855, Boston, MA, USA, 2015. IEEE. 3

[68] Thomas Pochart, Paulin Jacquot, and Joseph Mikael. On the challenges of using D-Wave computers to sample Boltzmann Random Variables, 2021. 2, 4, 5

[69] Andrew Rosenberg and Julia Hirschberg. V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 410–420, Prague, Czech Republic, 2007. Association for Computational Linguistics. 6

[70] E. Schrödinger. Discussion of Probability Relations between Separated Systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4):555–563, 1935. 3

[71] Peter W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Review*, 41(2):303–332, 1999. 3

[72] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, 2014. 8

[73] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, Cambridge, 2 edition, 2017. 3

[74] K. Wright, K. M. Beck, S. Debnath, J. M. Amini, Y. Nam, N. Grzesiak, J.-S. Chen, N. C. Pisenti, M. Chmielewski, C. Collins, K. M. Hudek, J. Mizrahi, J. D. Wong-Campos, S. Allen, J. Apisdorf, P. Solomon, M. Williams, A. M. Ducore, A. Blinov, S. M. Kreikemeier, V. Chaplin, M. Keesan, C. Monroe, and J. Kim. Benchmarking an 11-qubit quantum computer. *Nature Communications*, 10(1):5464, 2019. 2

[75] Rui Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005. 1

[76] Jianwei Yang, Devi Parikh, and Dhruv Batra. Joint Unsupervised Learning of Deep Representations and Image Clusters. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5147–5156, Las Vegas, NV, USA, 2016. IEEE. 1

[77] Jan-Nico Zaech, Alexander Liniger, Martin Danelljan, Dengxin Dai, and Luc Van Gool. Adiabatic Quantum Computing for Multi Object Tracking. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8801–8812, New Orleans, LA, USA, 2022. IEEE. 2, 3, 5, 6

[78] Hongyuan Zha, Xiaofeng He, Chris Ding, Ming Gu, and Horst Simon. Spectral relaxation for k-means clustering. *Advances in neural information processing systems*, 14, 2001. 1

[79] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint Discriminative and Generative Learning for Person Re-Identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2133–2142, Long Beach, CA, USA, 2019. IEEE. 3

Supplementary Material: Probabilistic Sampling of Balanced K-Means using Adiabatic Quantum Computing

Jan-Nico Zaech [1,2] Martin Danelljan [1] Tolga Birdal [3] Luc Van Gool [1,2]

[1]ETH Zurich, Switzerland [2] INSAIT, Sofia University, Bulgaria
[3]Imperial College London, United Kingdom

1. Introduction

The main manuscript presents a novel approach for probabilistic clustering based on exploiting the probabilistic nature of adiabatic quantum computing (AQC). In the supplementary material, we provide additional details and that complement the main manuscript. Section 2 presents the detailed derivation of the energy function used in the manuscript. Section 3 discusses the optimization of the inference parameters to increase the AQC solver performance. Section 4 provides information on data generation for the synthetic and real-world datasets used in the experiments. Sections 5, 6 and 7 extend the clustering performance and calibration evaluation on synthetic data and the IRIS dataset respectively. In Section 8, we discuss failure cases encountered during the experiments. Finally, we outline the limitations of our approach in Section 9.

Overall, the supplementary material aims to clarify open questions from the main manuscript, provides additional insights into the proposed method and discusses its current limitations.

2. Energy Function Derivation

The following section shows the step-by-step derivation of the energy function used in this paper. As clusters are independent, the energy for each cluster can be computed separately $E(X|Z) = \sum_k E_k(X|Z)$, where X represents the data-points and Z is the assignment matrix with entry Z_{ki} assigning point x_i to cluster c_k. For a single cluster c_k, the energy can be further extended into the quadratic and linear terms as follows

$$\begin{aligned} E_k(X|Z) &= \sum_i Z_{ki}(x_i - \mu_k)^\intercal I(x_i - \mu_k) \\ &= \sum_i Z_{ki}(x_i^\intercal I x_i - 2x_i^\intercal I\mu_k + \mu_k^\intercal I\mu_k) \\ &= \sum_i Z_{ki}x_i^\intercal I x_i - 2\sum_i Z_{ki}x_i^\intercal I\mu_k + \sum_i Z_{ki}\mu_k^\intercal I\mu_k \\ &= \sum_i Z_{ki}x_i^\intercal I x_i - 2\sum_i Z_{ki}x_i^\intercal I\mu_k + s_k\mu_k^\intercal I\mu_k, \end{aligned} \tag{1}$$

with the cluster mean μ_k and identity matrix I. By using the maximum likelihood (ML) estimator of the cluster mean

$$\mu_k = \frac{1}{s_k}\sum_j Z_{kj}x_j, \tag{2}$$

with cluster size s_k, the energy formulation only depends on the data and the cluster assignment

$$\begin{aligned} &\sum_i Z_{ki}x_i^\intercal I x_i - 2\sum_i Z_{ki}x_i^\intercal I\mu_k + s_k\mu_k^\intercal I\mu_k \\ &= \sum_i Z_{ki}x_i^\intercal I x_i - 2\sum_i Z_{ki}x_i^\intercal I\frac{1}{s_k}\sum_j Z_{kj}x_j \\ &+ s_k\frac{1}{s_k}\sum_i Z_{ki}x_i^\intercal I\frac{1}{s_k}\sum_j Z_{kj}x_j \\ &= \sum_i Z_{ki}x_i^\intercal I x_i - \frac{1}{s_k}\sum_i\sum_j Z_{ki}Z_{kj}x_i^\intercal I x_j. \end{aligned} \tag{3}$$

This finally shows that the total energy only depends on the distance between each pair of points

$$
\begin{aligned}
&\sum_i Z_{ki} x_i^\intercal I x_i - \frac{1}{s_k} \sum_i \sum_j Z_{ki} Z_{kj} x_i^\intercal I x_j \\
&= \frac{1}{s_k} \sum_i \sum_j Z_{ki} Z_{kj}\ x_i^\intercal I x_i - x_i^\intercal I x_j \\
&= \frac{1}{s_k} \sum_i \sum_j Z_{ki} Z_{kj}\ x_i^\intercal I (x_i - x_j) \\
&= \frac{1}{s_k} \sum_i \sum_j Z_{ki} Z_{kj}\ \frac{1}{2} [x_i^\intercal I (x_i - x_j) + x_j^\intercal I (x_j - x_i)] \\
&= \frac{1}{s_k} \sum_i \sum_j Z_{ki} Z_{kj}\ (x_i - x_j)^\intercal I (x_i - x_j).
\end{aligned}
\tag{4}
$$

3. Inference Parameter Optimization

Using the quadratic penalty method to include constraints in the Quadratic Binary Optimization (QUBO) formulation requires finding suitable Lagrangian multipliers λ. Even though a very high multiplier theoretically guarantees to find a feasible solution, it also deteriorates the conditioning of the optimization problem. Therefore, a suitable Lagrangian multiplier lifts the cost of any constraint violation above all relevant solutions of the clustering problem, while keeping them low enough to avoid scaling the total energy of the problem up considerably. To estimate the multipliers for each constraint, we follow an iterative procedure.

In an initial step, balanced k-means[2] is used to find a feasible clustering solution. This is used to offset the distance terms for each point such that the total clustering solution has an energy of 0. For the next steps, the constraints are separated into 3 components:

The cluster size constraint is defined by $\sum_i Z_{ki} = s_k\ \forall k$. Due to its strong diagonal term in its quadratic form using Lagrange multipliers, it quickly degrades the energy scaling of the problem. The Lagrangian multiplier corresponding to this constraint is estimated from the maximum cost improvement that can be achieved by switching one point between clusters.

The constraint $\sum_k Z_{ki} = 1\ \forall i$, which ensures the matching of every point to exactly one cluster, is further segmented into two parts. One part contains the positive off-diagonal elements and penalizes assigning a single point to multiple clusters. Its corresponding Lagrangian is computed from the maximum cost improvement that can be achieved by assigning an additional point to any cluster, compared to the k-means solution. The other term contains negative diagonal elements and adds an incentive to assign every point to one cluster. The corresponding multiplier is estimated by the maximum cost improvement by removing one point from a cluster and thus violating the constraint.

In five subsequent optimization steps the clustering problem is solved using simulated annealing and the Lagrangian multipliers are increased for constraints that are not fulfilled.

In the last step, which is only performed for simulated annealing, measurements at low temperatures of the Boltzmann distribution are handled. In scenarios where only a single valid clustering solution is returned, the Lagrangian multiplier of the cluster size constraint is increased, which results in sampling the problem at a higher temperature.

Finding well-suited Lagrangian multipliers is crucial due to the low fidelity of the current generation of AQCs, which requires careful engineering of the problem energy. Therefore, we expect this procedure to become of reduced importance in future generations of lower noise AQCs.

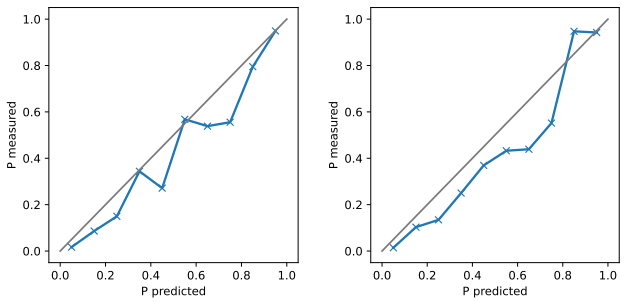
For experiments performed on the D-Wave AQC, all Lagrangian optimization steps are performed with SIM, before measuring the final results on the AQC due to the strong compute-time limitations.

4. Data Generation

Synthetic Data is generated by sampling a total of I points from separate normal distributions for each of K clusters. The cluster centers are selected as the corners of a simplex with uniformly drawn edge length, such that the distance between each pair of clusters lies within a predefined range $[d_{\text{min}}, d_{\text{max}}]$. The feature-space needs to be at least $K-1$ dimensional for each clustering problem. Sampling the edge-length randomly allows to generate a wide range of clustering problems with a different degree of ambiguity, due to the changing degree of overlap between distributions. This allows to evaluate the whole range of predicted posterior probability values. For each experiment a total of L clustering tasks is generated to evaluate the clustering metrics.

IRIS [1] is subsampled to generate quantitative results over different clustering scenarios. The whole IRIS dataset contains 3 classes, 50 samples for each class and 4 features forming a 4-dimensional space. According to the experiment parameters we randomly select a subset of classes, samples and features without replacement to allow running the tasks on a D-wave AQC. This generates different clustering problems, while keeping the general structure of the data in IRIS.

Image data is used to demonstrate the applicability of our method to computer vision tasks. We collected 8 images each for cars and boats and two images of cars towing boats. Visual embeddings for each image are extracted after the last layer of VGG16 [3] pretrained on Imagenet. Subsequently, the high-dimensional features are clustered using our approach, and the first coreset is computed to identify the ambiguous samples as demonstrated in the main paper.



(a) 3 clusters, 30 points. (b) 3 clusters, 45 points.

Figure 1. Evaluation of the calibration for simulated annealing in clustering scenarios with 3 clusters and 30/45 points respectively. All results generated with 1,000 problems in each scenario and 20,000 measurements for each clustering problem.

5. Cluster Calibration Evaluation

This section extends the analysis of the calibration of our method to additional synthetic scenarios with more variation and increased problem size. The plots provided in this section are generated similarly to the main manuscript where first all clustering solutions Z are accumulated in bins according to their estimated posterior probability $P(Z|X)$, including all sampled but non-optimal solutions. After accumulation, the ratio of correct solutions in each bin is evaluated and plotted over the probability range of each bin. In the plots the diagonal represents the desired calibration.

Further calibration plots for the scenario with 3 clusters and an increasing number of total points are provided in Figure 1. The scenarios are solved using simulated annealing with 20,000 measurements for each problem. The experiment with a total of 45 points in Figure 1b shows an overestimation of the posterior probability of the respective solutions. This can be attributed to two possible scenarios, where 1) the best solution is found, but not all relevant high-energy solutions are found during annealing and 2) the lowest-energy solution is not found and thus, the probability of all other solutions is overestimated. As the optimization problem becomes harder with an increasing number of points, the behavior is stronger in Figure 1b than in Figure 1a.

6. Coreset Sparsification Performance

The set of feasible solutions can be merged by using the calibrated confidence scores in Algorithm 1 introduced in the main manuscript. It sequentially removes uncertain points from the solution thus, increasing the solution probability. In Figure 2, we show the probability of the best merged solution being correctly evaluated over the minimum solution probability of the sparsified coreset. It shows that our approach of removing single points can considerably increase

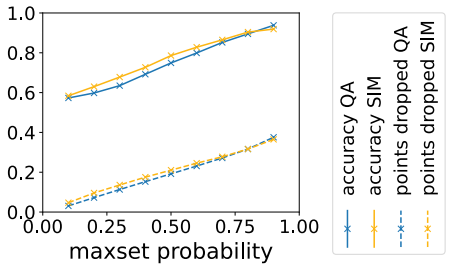


Figure 2. Clustering accuracy wrt. removing uncertain points by merging coresets.

the solution probability, thus highlighting the quality of the found coresets.

7. Evaluation on IRIS

Table 2 in the main manuscript provides performance metrics for randomly subsampled versions of the IRIS dataset [1]. In this section, we further evaluate the performance on the whole IRIS dataset, which contains 3 classes, 50 samples for each class and 4 features. We use simulated annealing with 20,000 measurements and balanced k-means to solve the IRIS clustering task, which both provide the same solution. The qualitative results are depicted in Figure 3, where all pairs of features are visualized. The shape of each sample represents the ground truth class and the color the result of the clustering algorithm. While the different feature pairs are plotted separately, the problem is solved as a single 4-dimensional clustering task. As results are identical with simulated annealing and balanced k-means, clustering metrics are also identical with a Completeness of 77.7%, Adjusted Rand index 78.6% and Fowlkes-Mallows Score of 85.6%.

8. Failure Cases

Analyzing the failure cases of our method provides valuable insight into the current state of quantum computing in computer vision, which aids to identify areas that need to be further investigated.

8.1. K-means

The analysis of synthetic problems in Table 1 in the main manuscript shows an advantage of our approach compared to the balanced k-means algorithm [2] for the smaller clustering scenarios. These cases can be traced back to the k-means algorithm finding local minima where switching points given the last cluster means does not improve the data fit. This scenario is avoided in our formulation by jointly optimizing for the assignment and cluster centers. Two examples of such failure cases of k-means clustering are provided in Figure 4.

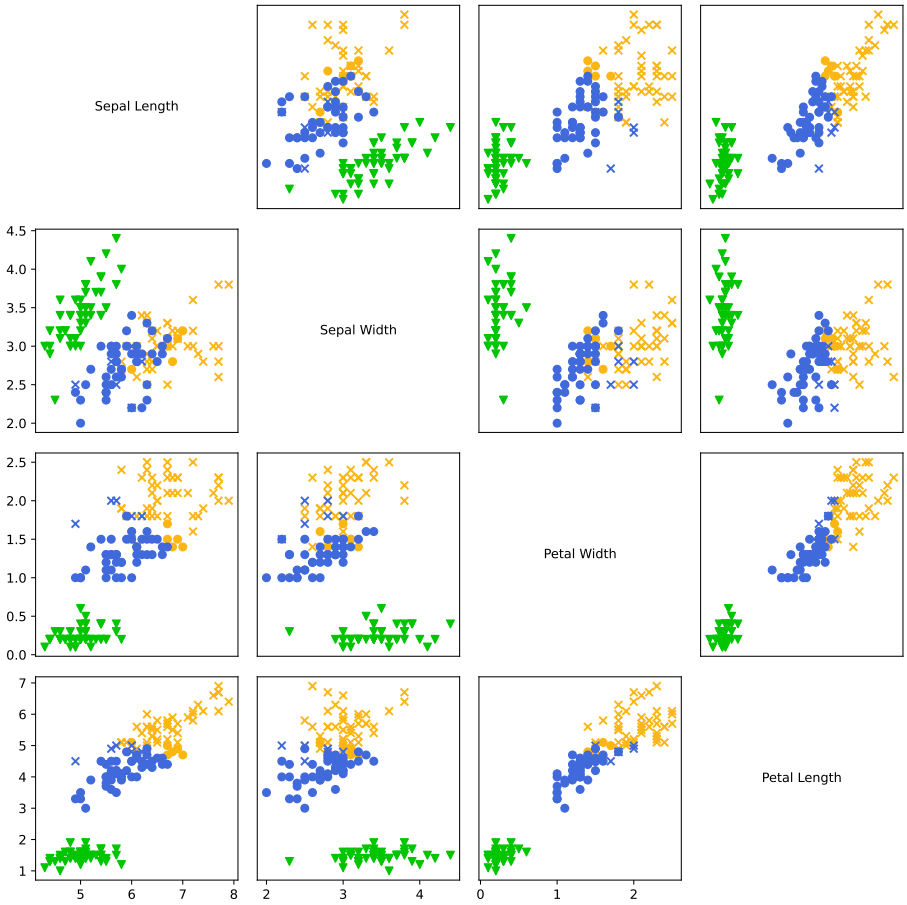


Figure 3. Clustering results on the IRIS dataset for simulated annealing and k-means.

8.1.1 Annealing based clustering

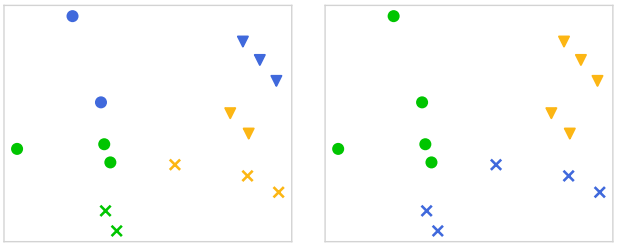
The scenario with a total of 45 Points in 3 Clusters shows an advantage of k-means in the number of correctly solved problems compared to our approach using simulated annealing. The main source for this behavior lies in not finding the lowest energy and thus, the optimal solution of the clustering problem, as depicted in Figures 5a and 5b. Another source of error in this scenario is shown if Figures 5c and 5d, where the local k-means solution corresponds to the ground truth, even though it has a higher energy. As the solutions returned by our approach are still dense clusters, the clustering metrics remain competitive with the balanced k-means approach.

9. Limitations

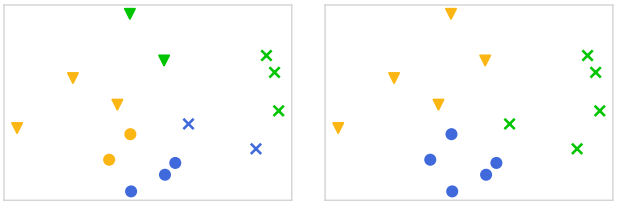
Our work aims at demonstrating the potential of using a quantum computer as a sampler for k-means clustering, in order to find multiple likely solutions and their associated calibrated posterior probabilities. Given the novelty of applying quantum computing to computer vision, it's natural that many works in this area, including ours, still come with limitations.

Current quantum computers are still limited in their fidelity of qubit couplings, which represent the terms of the quadratic cost function. This requires a careful selection of Lagrangian multipliers, which adds additional computational cost in the current formulation. With improving AQCs, this problem can be reduced and help to increase the problem size, as well as the robustness of the formulation. Another hardware limitation is the restricted connectivity between qubits. In the D-Wave Advantage 2 prototype used in this work, each qubit is coupled to up to 20 neighbors. This requires to build chains of qubits to represent a dense cost-matrix. Therefore, investigating sparse representations for clustering that reduce the required chain length can help to embed larger problems on the AQC.

Finally, our clustering approach is following the k-means cost function, with an identity covariance matrix. While this can model a range of practical problems, where the distribution of the data can directly be influenced, future work should investigate formulations of higher-order terms.

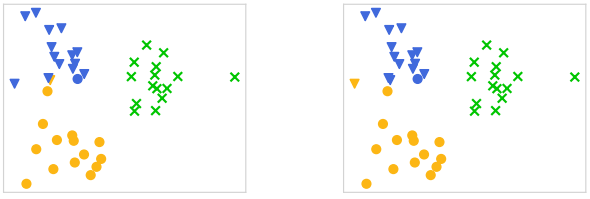


(a) K-means. (b) Simulated annealing.

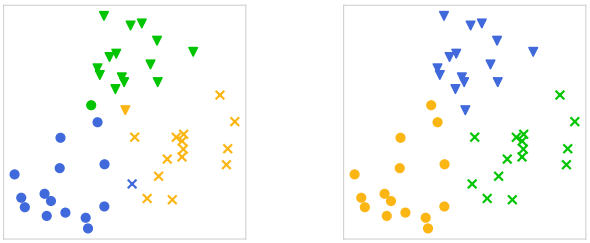


(c) K-means. (d) Simulated annealing.

Figure 4. Failure cases for k-means clustering. While our formulation finds the correct solution, k-means returns a local minimum.



(a) Simulated annealing. (b) K-means.



(c) Simulated annealing. (d) K-means.

Figure 5. Failure cases for simulated annealing clustering.

References

[1] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2):179–188, 1936. 2, 3

[2] Mikko I. Malinen and Pasi Fränti. Balanced K-Means for Clustering. In *Structural, Syntactic, and Statistical Pattern Recognition*, pages 32–41. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. 2, 3

[3] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, 2014. 2