The Method of Quantum Clustering

David Horn and Assaf Gottlieb
School of Physics and Astronomy
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv 69978, Israel

Abstract

We propose a novel clustering method that is an extension of ideas inherent to scale-space clustering and support-vector clustering. Like the latter, it associates every data point with a vector in Hilbert space, and like the former it puts emphasis on their total sum, that is equal to the scale-space probability function. The novelty of our approach is the study of an operator in Hilbert space, represented by the Schrödinger equation of which the probability function is a solution. This Schrödinger equation contains a potential function that can be derived analytically from the probability function. We associate minima of the potential with cluster centers. The method has one variable parameter, the scale of its Gaussian kernel. We demonstrate its applicability on known data sets. By limiting the evaluation of the Schrödinger potential to the locations of data points, we can apply this method to problems in high dimensions.

1 Introduction

Methods of data clustering are usually based on geometric or probabilistic considerations [1, 2, 3]. The problem of unsupervised learning of clusters based on locations of points in data-space, is in general ill defined. Hence intuition based on other fields of study may be useful in formulating new heuristic procedures. The example of [4] shows how intuition derived from statistical mechanics leads to successful results. Here we propose a model based on tools that are borrowed from quantum mechanics.

We start out with the scale-space algorithm of [5] that uses a Parzen-window estimator of the probability distribution based on the data. Using a Gaussian kernel, one generates from the $N$ data points in a Euclidean space of dimension $d$ a probability distribution given by,

$$
\psi(\mathbf{x}) = \sum_i e^{-\frac{|x-x_i|^2}{2\sigma^2}}
$$

(1)

where $x_i$ are the data points. It seems quite natural [5] to associate maxima of this function with cluster centers.

The same kind of Gaussian kernel was the basis of another method, Support Vector Clustering (SVC) [6], associating the $N$ data-points $x_i$ with vectors in an abstract Hilbert space.
Here we will also consider a Hilbert space, but, in contradistinction with kernel methods where the Hilbert space is implicit, here we work with a Schrödinger equation that serves as the basic framework of the Hilbert space. Our method was introduced in [7] and is further expanded in this presentation. Its main emphasis is on the Schrödinger potential, whose minima will determine the cluster centers. This potential is part of the Schrödinger equation that \( \psi \) is a solution of.

## 2 The Schrödinger Potential

We define[7] the Schrödinger equation

\[
H\psi = (-\frac{\sigma^2}{2} \nabla^2 + V(\mathbf{x}))\psi(\mathbf{x}) = E\psi(\mathbf{x})
\]  

(2)

for which \( \psi(\mathbf{x}) \) is a solution, or eigenstate. The simplest case is that of a single Gaussian, when \( \psi \) represents a single point at \( \mathbf{x}_1 \). Then it turns out that \( V = \frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{x}_1)^2 \). This quadratic function, whose center lies at \( \mathbf{x}_1 \), is known as the harmonic potential in quantum mechanics (see, e.g., [8]). Its eigenvalue \( E = d/2 \) is the lowest possible eigenvalue of \( H \), hence the Gaussian function is said to describe the ground state of \( H \).

Conventionally, in quantum mechanics, one is given \( H \) and one searches for solutions, or eigenfunctions, \( \psi(\mathbf{x}) \). Here, we have already \( \psi(\mathbf{x}) \), as determined by the data points, we ask therefore for the \( V(\mathbf{x}) \) whose solution is the given \( \psi(\mathbf{x}) \). This can be easily obtained through

\[
V(\mathbf{x}) = E + \frac{\sigma^2}{\psi} \nabla^2 \psi = \frac{d}{2} + \frac{1}{2\sigma^2} \sum_i (\mathbf{x} - \mathbf{x}_i)^2 e^{-\frac{\mathbf{x}_i^2}{2\sigma^2}}
\]  

(3)

\( E \) is still left undefined. For this purpose we require \( V \) to be positive definite, i.e. \( \min V = 0 \). This sets the value of

\[
E = -\min \frac{\sigma^2}{\psi} \nabla^2 \psi
\]  

(4)

and determines \( V(\mathbf{x}) \) uniquely. Using Eq. 3 it is easy to prove that

\[
0 < E \leq \frac{d}{2}
\]  

(5)

## 3 2D Examples

### 3.1 Crab Data

To show the power of our new method we discuss the crab data set taken from Ripley’s book [9]. This data set is defined over a five-dimensional parameter space. When analyzed in terms of the 2nd and 3rd principal components of the correlation matrix one observes a nice separation of the 200 instances into their four classes. We start therefore with this problem as our first test case. In Fig. 1 we show the data as well as the Parzen probability distribution \( \psi(\mathbf{x}) \) using the width parameter \( \sigma = 1/\sqrt{2} \). It is quite obvious that this width is not small enough to deduce the correct clustering according to the approach of [5]. Nonetheless, the potential displayed in Fig. 2 shows the required four minima for the same width parameter. Thus we conclude that the necessary information is already available. One needs, however, the quantum clustering approach, to bring it out.

\( H \) (the Hamiltonian) and \( V \) (potential energy) are conventional quantum mechanical operators, rescaled so that \( H \) depends on one parameter, \( \sigma \). \( E \) is a (rescaled) energy eigenvalue in quantum mechanics.