Quantum-like Gaussian mixture model

Andreas Wichert

Accepted: 3 June 2021 / Published online: 11 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
A new concept of a quantum-like mixture model is introduced. It describes the mixture distribution with the assumption that a point is generated by each Gaussian at the same time. The quantum-like mixture Gaussian improves the classification accuracy in machine learning by indicating that the uncertain points should not be assigned to any class. It increases the accuracy of the mixture Gaussian model on the iris data set from 96.67 to 99.24%.

Keywords
Clustering · Quantum probabilities · Interference · Gaussian mixture model

1 Introduction

We introduce a new concept of a quantum-like mixture model inspired by quantum physics. This concept describes the mixture distribution with the assumption that a point is generated by each Gaussian at the same time. Between different Gaussians interference is present that changes the shape of the mixture distribution. The shape of the distributions is changed in such a way that the probability of points between the Gaussians is lowered and in the center of the Gaussians is increased. A difference between the relative probabilities determined by the mixture Gaussians and the relative probabilities that are defined by quantum-like mixture Gaussians indicates the overlapping error that results from the striking distance to a decision boundary. The quantum-like mixture Gaussian improves the classification quality of a Gaussian classifier. Points that cannot be classified correctly, because they are too close to a decision boundary, are not considered.

1.1 Quantum probabilities

In quantum physics a probability \( p(x) \) of a state \( x \) as the squared magnitude of a probability amplitude \( A(x) \) is evaluated, which is represented by a complex number

\[
p(x) = |A(x)|^2 = A(x)^* \cdot A(x).
\]  

This is because the product of complex number with its conjugate is always a real number. With

\[
A(x) = \alpha + \beta \cdot i
\]

\[
A(x)^* \cdot A(x) = (\alpha - \beta \cdot i) \cdot (\alpha + \beta \cdot i)
\]

\[
A(x)^* \cdot A(x) = \alpha^2 + \beta^2 = |A(x)|^2.
\]

Quantum physics by itself does not offer any justification or explanation beside the statement that it just works fine, see Binney and Skinner (2014). These quantum probabilities are as well called von Neumann probabilities.

1.1.1 Two-slit interference

Suppose there are two mutual exclusive events \( x \) and \( y \). This means \( x \) and \( y \) do not occur together Wichert (2020). The classical probability of an event \( x \) or event \( y \) is just

\[
p(x \lor y) = p(x) + p(y).
\]  

This is the sum rule for probabilities for exclusive events. For probability amplitudes, it is as well

\[
A(x \lor y) = A(x) + A(y).
\]
However, converting these amplitudes into probabilities according to Eq. 1 leads to an interference term $2 \cdot \Re(A(x) \cdot A^*(y))$,

$|A(x) + A(y)|^2 = |A(x)|^2 + |A(y)|^2 + 2 \cdot \Re(A(x) \cdot A^*(y))$

$|A(x) + A(y)|^2 = p(x) + p(y) + 2 \cdot \Re(A(x) \cdot A^*(y)),$ \hspace{1cm} (7)

making both approaches, in general, incompatible

$|A(x) + A(y)|^2 \neq p(x) + p(y).$ \hspace{1cm} (8)

In other words, the summation rule of classical probability theory is violated, resulting in one of the most fundamental laws of quantum mechanics, see Binney and Skinner (2014). Suppose that an unobserved state evolves smoothly and continuously. However, during the measurement, it collapses into a definite state with a probability $p(x) = |A(x)|^2$. For instance, let us imagine a gun that fires electrons and a screen with two narrow slits $x$ and $y$ and a photographic plate. An emitted electron can pass through slit $x$ or slit $y$ and reaches the photographic plate at the position $z$, which is equidistant from both slits. The electron detectors show from which slit the electron went through, and we find that the probability of the electron hitting the photographic plate is

$p(z) = p(x) + p(y).$ \hspace{1cm} (9)

This probability means that, when measured, the electron behaved as a particle. On the other hand, if we remove the detectors, the electron is unobserved, not knowing through which slit it went through. Now, the electron is represented as a wave with the amplitudes

$a(x, \theta_1) = \sqrt{p(x)} \cdot e^{i \theta_1} = A(x),$ \hspace{1cm} (10)

$a(y, \theta_2) = \sqrt{p(y)} \cdot e^{i \theta_2} = A(y).$ \hspace{1cm} (11)

These amplitudes contain a parameter $\theta$, which corresponds to the phase of the wave. The equation then becomes

$|a(x, \theta_1) + a(y, \theta_2)|^2 = (a(x, \theta_1) + a(y, \theta_2)) \cdot (a(x, \theta_1) + a(y, \theta_2))^*$ \hspace{1cm} (12)

with

$p(z) \neq |a(x, \theta_1) + a(y, \theta_2)|^2 \geq 0$

for most values of $\theta_1, \theta_2$. Since the norm is being positive or more precisely non-negative, the value are always bigger equal zero. Then,

$|a(x, \theta_1) + a(y, \theta_2)|^2 = \left(\sqrt{p(x)} \cdot e^{i \theta_1} + \sqrt{p(y)} \cdot e^{i \theta_2}\right) \cdot \left(\sqrt{p(x)} \cdot e^{-i \theta_1} + \sqrt{p(y)} \cdot e^{-i \theta_2}\right) = p(x) + p(y) + \sqrt{p(y)} \cdot p(x) \cdot e^{i(\theta_2 - \theta_1)} + \sqrt{p(x)} \cdot p(y) \cdot e^{i(\theta_1 - \theta_2)}$

$= p(x) + p(y) + \sqrt{p(x) \cdot p(y)} \cdot e^{i(\theta_1 - \theta_2)} \cdot \left(e^{i(\theta_1 - \theta_2)} + e^{-i(\theta_1 - \theta_2)}\right).$ \hspace{1cm} (13)

With

$\cos(\theta_1 - \theta_2) = \frac{e^{i(\theta_1 - \theta_2)} + e^{-i(\theta_1 - \theta_2)}}{2},$ \hspace{1cm} (14)

we simplify to

$p(x) + p(y) + 2 \cdot \sqrt{p(x) \cdot p(y)} \cdot \cos(\theta_1 - \theta_2).$ \hspace{1cm} (15)

The amplitude is the root of the belief multiplied with the corresponding phase with the phase $\theta \in [0, 2 \cdot \pi)$

$A(x) = a(x, \theta) = \sqrt{p(x)} \cdot e^{i \theta}.$ \hspace{1cm} (16)

Empirical evidence from the cognitive psychology literature suggests that most human decision-making cannot be adequately modeled using classical probability theory as defined by Kolmogorov’s axioms (Tversky and Kahneman 1973; Tversky and Kahneman 1974). These empirical findings show that, under uncertainty, humans tend to violate the expected utility theory and consequently the laws of classical probability theory [e.g., the law of total probability (Busemeyer et al. 2006, 2009; Khrennikov 2009)], Busemeyer et al. (2011), Yukalov and Sornette (2011), Busemeyer et al. (2012), Busemeyer and Wang (2014), (Wichert and Moreira 2018; Wichert et al. 2020), Wichert (2020) leading to what is known as the “disjunction effect” which, in turn, leads to violation of the Sure Thing Principle. The violation results from an additional interference that influences the classical probabilities. We present new findings in which the interference is present in the quantum-like mixture of Gaussians. They describes the mixture distribution with the assumption that a point is generated by each Gaussian at the same time. The decision boundary of a quantum-like mixture Gaussian separates probabilities. This leads to a better interpretation of the switching Kalman filter and to an improvement of the classification accuracy in quantum-like machine learning Sergioli (2020), Sergioli et al. (2021).
2 Quantum-like mixture of Gaussians

The Gaussian distribution or normal distribution is defined as a

\[
\rho(x|\mu, \sigma^2) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x - \mu)^2}{\sigma^2}\right)
\]

(17)

where \(\mu\) is the mean, \(\sigma\) is the standard deviation and \(\sigma^2\) is the variance and \(\rho(x|\mu, \sigma^2)\) represents the conditional relative probability. The Gaussian distribution or normal distribution is defined as a probability density function (PDF) that reflects the relative probability. The PDF can be bigger one for small \(\sigma\) since it is the area under the PDF that represents the probability. However, the PDF reflects the relative probability \(\rho\) (rho) and in the following we will write \(\rho(x|\mu, \sigma^2)\) to indicate the relative probability.

The Gaussian distribution can be generalized to a \(D\)-dimensional space with

\[
\rho(x|\mu, \Sigma) = \mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2}\cdot(x - \mu)^T \Sigma^{-1} \cdot (x - \mu)\right)
\]

(18)

where \(\mu\) is the \(D\)-dimensional mean vector, \(\Sigma\) is a \(D \times D\) covariance matrix and \(|\Sigma|\) is the determinant of \(\Sigma\).

A Gaussian mixture distribution is a combination of multiple Gaussians where each has a certain weight

\[
\rho(x) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x|\mu_k, \Sigma_k)
\]

(19)

\[
\rho(x) = \sum_{k=1}^{K} \pi_k \cdot \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \cdot \exp\left(-\frac{1}{2}\cdot(x - \mu_k)^T \Sigma_k^{-1} \cdot (x - \mu_k)\right)
\]

(20)

with

\[0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1.\]

2.1 Gaussian amplitudes

We will investigate a simple case of Gaussian mixture model with

\[
\rho(x) = \pi_1 \cdot \rho(x|\mu_1, \sigma_1^2) + \pi_2 \cdot \rho(x|\mu_2, \sigma_2^2)
\]

(21)

\[
\rho(x) = \pi_1 \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x - \mu_1)^2}{\sigma_1^2}\right) + \pi_2 \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x - \mu_2)^2}{\sigma_2^2}\right)
\]

(22)

See in Fig. 1 a simple example.

With the Gaussian relative amplitude \(\alpha\) with parameter \(\theta\)

\[
\rho(x|\mu_1, \sigma_1^2) = \alpha(x, \theta_1|\mu_1, \sigma_1^2) \cdot \alpha(x, \theta_1|\mu_1, \sigma_1^2)
\]

\[
\alpha(x, \theta_1|\mu_1, \sigma_1^2) = \frac{1}{(2\pi\sigma_1^2)^{1/2}} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x - \mu_1)^2}{\sigma_1^2}\right) \cdot e^{\theta_1}
\]

(23)

\[
\alpha(x, \theta_2|\mu_2, \sigma_2^2) = \frac{1}{(2\pi\sigma_2^2)^{1/2}} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x - \mu_2)^2}{\sigma_2^2}\right) \cdot e^{\theta_2}
\]

(24)

A quantum-like Gaussian mixture model is defined as

\[
|\sqrt{\pi_1} \cdot \alpha(x, \theta_1|\mu_1, \sigma_1^2) + \sqrt{\pi_2} \cdot \alpha(x, \theta_2|\mu_2, \sigma_2^2)|^2
\]

\[
= (|\sqrt{\pi_1} \cdot \alpha(x, \theta_1|\mu_1, \sigma_1^2) + \sqrt{\pi_2} \cdot \alpha(x, \theta_2|\mu_2, \sigma_2^2)|)^2
\]

\[
\cdot (|\sqrt{\pi_1} \cdot \alpha(x, \theta_1|\mu_1, \sigma_1^2) + \sqrt{\pi_2} \cdot \alpha(x, \theta_2|\mu_2, \sigma_2^2)|)^* =
\]

\[
\pi_1 \cdot \rho(x|\mu_1, \sigma_1^2) + \pi_2 \cdot \rho(x|\mu_2, \sigma_2^2)
\]

\[
+ 2 \cdot \sqrt{\pi_1 \cdot \pi_2} \cdot \sqrt{\rho(x|\mu_1, \sigma_1^2) \cdot \rho(x|\mu_2, \sigma_2^2)} \cdot \cos(\theta_1 - \theta_2).
\]

(25)

2.1.1 Multiplication of Gaussian amplitudes

For multiplication of Gaussians by two, it follows

\[
2 \cdot \rho(x|\mu, \sigma^2) = 0.5 \cdot \rho(x|\mu, \sigma^2) + 0.5 \cdot \rho(x|\mu, \sigma^2)
\]

\[
= \rho(x|\mu, \sigma^2).
\]

(27)

In the next step, we interpret the multiplication of Gaussian amplitudes \(\sqrt{\rho(x|\mu_1, \sigma_1^2)} \cdot \sqrt{\rho(x|\mu_2, \sigma_2^2)}\). We will follow the approach as described in Freier (2013). Given

\[
\frac{1}{(2\pi\sigma_1^2)^{1/2}} \cdot e^{-\frac{(x - \mu_1)^2}{4\sigma_1^2}} \cdot \frac{1}{(2\pi\sigma_2^2)^{1/2}} \cdot e^{-\frac{(x - \mu_2)^2}{4\sigma_2^2}}
\]

\[
= c \cdot \frac{1}{(2\pi\sigma_1^2)^{1/2}} \cdot e^{-\frac{(x - \mu_1)^2}{4\sigma_1^2}} \cdot \frac{1}{(2\pi\sigma_2^2)^{1/2}} \cdot e^{\theta_1}
\]

\[
\cdot \frac{1}{(2\pi\sigma_2^2)^{1/2}} \cdot e^{\theta_2} = c_1 \cdot e^{\theta_2} \cdot \frac{1}{(2\pi\sigma_2^2)^{1/2}} \cdot e^{\theta_2}
\]

it follows

\[
\rho(x) = \pi_1 \cdot \rho(x|\mu_1, \sigma_1^2) + \pi_2 \cdot \rho(x|\mu_2, \sigma_2^2)
\]

\[
\rho(x) = \pi_1 \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x - \mu_1)^2}{\sigma_1^2}\right) + \pi_2 \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \cdot \exp\left(-\frac{1}{2}\cdot\frac{(x - \mu_2)^2}{\sigma_2^2}\right)
\]
Two Gaussians with $\mu_1 = 3, \sigma_1 = 3$ and $\mu_2 = -3, \sigma_2 = 2$. 

Gaussian mixture of the two Gaussians with equal mixture coefficients $\pi$

\[ e^{\alpha_1 + \alpha_2} = e^{\alpha_1 + \alpha_2 - \frac{(x - \mu_1)^2}{4 \cdot \sigma_1^2} - \frac{(x - \mu_2)^2}{4 \cdot \sigma_2^2}} = c_2 - \frac{(x - \mu)^2}{4 \cdot \sigma^2} \]

\[-4 \cdot c_2 + \frac{(x - \mu)^2}{\sigma^2} = \frac{(x - \mu_1)^2}{\sigma_1^2} + \frac{(x - \mu_2)^2}{\sigma_2^2} - 4 \cdot c_2 + \frac{(x - \mu)^2}{\sigma^2} = \frac{x^2 - 2 \cdot x \cdot \mu_1 + \mu_1^2}{\sigma_1^2} \]

\[+ \frac{x^2 - 2 \cdot x \cdot \mu_2 + \mu_2^2}{\sigma_2^2} - 4 \cdot c_2 + \frac{(x - \mu)^2}{\sigma^2} = \frac{\sigma_1^2 \cdot x - 2 \cdot \sigma_2^2 \cdot x \cdot \mu_1 + \sigma_2^2 \cdot \mu_2^2 + \sigma_1^2 \cdot x^2 - 2 \cdot \sigma_1^2 \cdot x \cdot \mu_2 + \sigma_1^2 \cdot \mu_2^2}{\sigma_1^2 \cdot \sigma_2^2} \]

\[-4 \cdot c_2 + \frac{(x - \mu)^2}{\sigma^2} = \frac{(\sigma_2^2 + \sigma_1^2) \cdot x^2 - 2 \cdot x \cdot (\sigma_2^2 \cdot \mu_1 + \sigma_1^2 \cdot \mu_2) + \sigma_2^2 \cdot \mu_1^2 + \sigma_1^2 \cdot \mu_2^2}{\sigma_1^2 \cdot \sigma_2^2} \]

\[-4 \cdot c_2 + \frac{x^2 - 2 \cdot x \cdot \mu + \mu^2}{\sigma^2} = \frac{x^2 - 2 \cdot x \cdot \frac{\sigma_2^2 \cdot \mu_1 + \sigma_1^2 \cdot \mu_2}{\sigma_2^2 + \sigma_1^2} + \frac{\sigma_2^2 \cdot \mu_1^2 + \sigma_1^2 \cdot \mu_2^2}{\sigma_2^2 + \sigma_1^2}}{\frac{\sigma_2^2 + \sigma_1^2}{\sigma_2^2 + \sigma_1^2}} \]

\[-4 \cdot c_2 + \frac{x^2 - 2 \cdot x \cdot \mu + \mu^2}{\sigma^2} = \frac{x^2 - 2 \cdot x \cdot \frac{\sigma_2^2 \cdot \mu_1 + \sigma_1^2 \cdot \mu_2}{\sigma_2^2 + \sigma_1^2} + \frac{\sigma_2^2 \cdot \mu_1^2 + \sigma_1^2 \cdot \mu_2^2}{\sigma_2^2 + \sigma_1^2}}{\frac{\sigma_2^2 + \sigma_1^2}{\sigma_2^2 + \sigma_1^2}} + \left(\frac{\sigma_2^2 \cdot \mu_1^2 + \sigma_1^2 \cdot \mu_2^2}{\sigma_2^2 + \sigma_1^2}\right)^2 \left(\frac{\sigma_1^2 \cdot \mu_1^2}{\sigma_2^2 + \sigma_1^2}\right)^2 \]

We assume:

\[ \frac{x^2 - 2 \cdot x \cdot \mu + \mu^2}{\sigma^2} = \frac{x^2 - 2 \cdot x \cdot \frac{\sigma_2^2 \cdot \mu_1 + \sigma_1^2 \cdot \mu_2}{\sigma_2^2 + \sigma_1^2} + \left(\frac{\sigma_2^2 \cdot \mu_1 + \sigma_1^2 \cdot \mu_2}{\sigma_2^2 + \sigma_1^2}\right)^2}{\frac{\sigma_2^2 + \sigma_1^2}{\sigma_2^2 + \sigma_1^2}} \]

\[ \sigma^2 = \frac{\sigma_1^2 \cdot \sigma_2^2}{\sigma_1^2 + \sigma_1^2} = \sigma_1^2 - \frac{\sigma_1^4}{\sigma_2^2 + \sigma_1^2} \]

and using the equality we get

\[ \mu = \frac{\sigma_2^2 \cdot \mu_1 + \sigma_1^2 \cdot \mu_2}{\sigma_2^2 + \sigma_1^2} = \mu_1 + \frac{\sigma_1^2 \cdot (\mu_2 - \mu_1)}{\sigma_2^2 + \sigma_1^2} \]

and it follows

\[ \mu = \frac{\sigma_2^2 \cdot \mu_1 + \sigma_1^2 \cdot \mu_2}{\sigma_2^2 + \sigma_1^2} = \mu_1 + \frac{\sigma_1^2 \cdot (\mu_2 - \mu_1)}{\sigma_2^2 + \sigma_1^2} \]
Now we determine \( c_1 \)

\[
\frac{1}{(2 \cdot \pi \cdot \sigma_1^2)^{\frac{1}{4}}}, e^{\theta_1}, \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}}, e^{\theta_2} = c_1 \cdot e^{\theta_2} \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}}, e^{\theta_2}.
\]

We know that

\[
\frac{1}{(2 \cdot \pi \cdot \sigma_1^2)^{\frac{1}{4}}} \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}} = c_1 \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}}
\]

and

\[
\frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_1 \cdot \sigma_2}} = c_1 \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}}
\]

\[
c_1 = \frac{1}{(2 \cdot \pi \cdot (\sigma_1^2 + \sigma_2^2))^{\frac{1}{4}}}.
\]

### 2.2 Interference waves

The quantum-like Gaussian mixture model is defined as

\[
\begin{align*}
|\sqrt{\pi_1} \cdot \alpha(x, \theta_1 | \mu_1, \sigma_1^2) + \sqrt{\pi_2} \cdot \alpha(y, \theta_2 | \mu_2, \sigma_2^2)|^2 \\
= \pi_1 \cdot \rho(x | \mu_1, \sigma_1^2) + \pi_2 \cdot \rho(x | \mu_2, \sigma_2^2) \\
+ \sqrt{\pi_1 \cdot \pi_2 \cdot c_1 \cdot e^{\theta_2} \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}} \cdot e^{\theta_2} \cdot \cos(\theta_1 - \theta_2)}
\end{align*}
\]

\[
\pi_1 \cdot \rho(x | \mu_1, \sigma_1^2) + \pi_2 \cdot \rho(x | \mu_2, \sigma_2^2)
+ \sqrt{\pi_1 \cdot \pi_2 \cdot c_1 \cdot e^{\theta_2} \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}} \cdot e^{\theta_2} \cdot \cos(\theta)}
\]

since we can simplify to \( \theta = \theta_1 - \theta_2 \) with

\[
\rho(x, \theta) = \pi_1 \cdot \rho(x | \mu_1, \sigma_1^2) + \pi_2 \cdot \rho(x | \mu_2, \sigma_2^2) \\
+ \sqrt{\pi_1 \cdot \pi_2 \cdot c_1 \cdot e^{\theta_2} \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}} \cdot e^{\theta_2} \cdot \cos(\theta)}
\]

The interference part value depends on \( \theta \), for \( \theta = 0 \) we have the maximal interference and for \( \theta = \pi \) the minimal interference. It means that

\[
\int_{-\infty}^{\infty} \rho(x, 0) \, dx > 1, \quad \int_{-\infty}^{\infty} \rho(x, \pi) \, dx < 1
\]

However it should be

\[
\int_{-\infty}^{\infty} \rho(x, \theta) \, dx = 1
\]

which is present with \( \theta = \pi/2 \) resulting in no interference, since \( \cos(\pi/2) = 0 \).

We can satisfy Eq. 38 for not fixed \( \theta \) values. By the linear change of \( \theta \), the interference part becomes a wave canceling it out so that Eq. 38 is valid. A wave is determined by the frequency \( f \) is the inverse of the period \( \rho \).

\[
f = \frac{1}{\rho}
\]

with the interference represented as a wave

\[
\sqrt{\pi_1 \cdot \pi_2 \cdot c_1 \cdot e^{\theta_2} \cdot \frac{1}{(2 \cdot \pi \cdot \sigma_2^2)^{\frac{1}{4}}} \cdot e^{\theta_2} \cdot \cos(f \cdot x)}
\]

For frequency \( f \rightarrow \infty \), the wave of Eq. 38 is satisfied with

\[
\lim_{f \to \infty} \int_{-\infty}^{\infty} \rho(x, x \cdot f) \, dx = 1
\]
In the case \( f \to \infty \), the relative probability \( \rho(x, x \cdot f) \) of \( x \) is either \( \rho(x, 0) \) or \( \rho(x, \pi) \), and we are uncertain about the value.

With a finite value \( f \), the value becomes certain. Figure 2 indicates an example with \( f = 10 \).

The shape of the Gaussian mixture changes in such a way that the two individual Gaussians become separated with a frequency defined by the two mean values

\[
f = \frac{2 \cdot \pi}{|\mu_1 - \mu_2|}
\]

and the phase by the smaller mean value \( \mu_{1,2}^{\text{min}} \)

\[
\mu_{1,2}^{\text{min}} = \begin{cases} 
\mu_1, & \text{if } \mu_1 \leq \mu_2 \\
\mu_2, & \text{otherwise}
\end{cases}
\]

with

\[
\cos(\theta) = \cos \left( \frac{(x - \mu_{1,2}^{\text{min}}) \cdot 2 \cdot \pi}{|\mu_1 - \mu_2|} \right)
\]

Figure 3 shows an example with \( f = 10 \).

2.2.1 Generalizations for several Gaussian

A quantum-like Gaussian mixture model is defined as

\[
\rho(x) = \sum_{k=1}^{K} \sqrt{\alpha_k} \cdot \alpha(x, \theta_k | \mu_k, \sigma_k^2) \]

\[
= \rho(x) + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \sqrt{\alpha_i \cdot \alpha_j} \cdot \sqrt{\rho(x | \mu_i, \sigma_i^2)} \\
\cdot \sqrt{\rho(x | \mu_j, \sigma_j^2)} \cdot \cos(\theta_i - \theta_j)
\]

Figures 3 and 4 show examples of quantum-like Gaussian mixture with small probability values between the two Gaussians.

\[
\int_{-\infty}^{\infty} \rho(x) \cdot \frac{(x - \mu_{1,2}^{\text{min}}) \cdot 2 \cdot \pi}{|\mu_1 - \mu_2|} \, dx \approx 1.
\]
With the Gaussian amplitude with parameter $\theta$.

$$
\rho(x | \mu_1, \Sigma_1) = \alpha(x, \theta_1 | \mu_1, \Sigma_1) \cdot \alpha(x, \theta_1 | \mu_1, \Sigma_1)
$$

$$
\rho(x, \theta_2 | \mu_2, \Sigma_2) = \frac{1}{(2 \cdot \pi)^{D/4} \cdot |\Sigma_1|^{1/4}} \cdot \frac{1}{|\Sigma_2|^{1/4}} \cdot \exp \left( -\frac{1}{4} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \right) \cdot e^{i \theta_1}
$$

$$
\rho(x, \theta_2 | \mu_2, \Sigma_2) = \frac{1}{(2 \cdot \pi)^{D/4} \cdot |\Sigma_1|^{1/4}} \cdot \frac{1}{|\Sigma_2|^{1/4}} \cdot \exp \left( -\frac{1}{4} (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \right) \cdot e^{i \theta_2}
$$

A quantum-like Gaussian mixture model is defined as

$$
|\sqrt{\pi_1} \cdot \alpha(x, \theta_1 | \mu_1, \Sigma_1) + \sqrt{\pi_2} \cdot \alpha(x, \theta_2 | \mu_2, \Sigma_2)|^2
$$

$$
= (\sqrt{\pi_1} \cdot \alpha(x, \theta_1 | \mu_1, \Sigma_1) + \sqrt{\pi_2} \cdot \alpha(x, \theta_2 | \mu_2, \Sigma_2)) \cdot (\sqrt{\pi_1} \cdot \alpha(x, \theta_1 | \mu_1, \Sigma_1) + \sqrt{\pi_2} \cdot \alpha(x, \theta_2 | \mu_2, \Sigma_2))
$$

$$
= \sqrt{\pi_1} \cdot \rho(x | \mu_1, \Sigma_1) + \rho(x | \mu_2, \Sigma_2)
$$

$$
+ \sqrt{\pi_1} \cdot \rho(x | \mu_1, \Sigma_1) \cdot \rho(x | \mu_2, \Sigma_2) \cdot \cos(\theta_1 - \theta_2).
$$

2.2.2 Generalization for $D$ dimensional space

We will follow the approach as described in Freier (2013).

$$
\mu = \mu_1 + \Sigma_1 \cdot (\Sigma_1 + \Sigma_2)^{-1} \cdot (\mu_2 - \mu_1),
$$

$$
\Sigma = \Sigma_1 - \Sigma_1 \cdot (\Sigma_1 + \Sigma_2)^{-1} \cdot \Sigma_1,
$$

The values of $c_2$ and $c_1$ are

$$
c_2 = -\frac{1}{4} \cdot (\mu_1 - \mu_2)^T \cdot (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2),
$$

$$
c_2 = \frac{(\mu_1 - \mu_2)^2}{4 \cdot (\sigma_2^2 + \sigma_1^2)}.
$$
Figures 6 and 7 show examples of quantum-like Gaussian mixture with small probability values between the two-dimensional Gaussians with \( \mu_1 = (0, -2)^T, \Sigma_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix} \) and \( \mu_2 = (-2, 0)^T, \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \) with equal mixture coefficients \( \pi \).

2.2.4 Generalizations for several Gaussian

A quantum-like Gaussian mixture model is defined as
Fig. 6  a Two Gaussians. b Gaussian mixture of the two Gaussians with equal mixture coefficients $\pi$. c Quantum-like Gaussian mixture with the interference wave with $\theta = \left(\frac{\langle x | \mu_i \rangle - \langle x | \mu_j \rangle}{\| \mu_i - \mu_j \|}\cdot 2 \cdot \pi \right)$ of the two Gaussians with equal mixture coefficients $\pi$

$$
\sum_{k=1}^{K} \sqrt{\pi_k} \cdot \alpha (x, \theta_k | \mu_k, \Sigma_k) \\
= \rho (x) + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \sqrt{\pi_i \cdot \pi_j} \\
\cdot \sqrt{p (x | \mu_i, \Sigma_i)} \cdot \rho (x | \mu_j, \Sigma_j) \cdot \cos (\theta_i - \theta_j)$$

(56)

with the normed direction vector

$$
\mu_{i,j} = \frac{\langle \mu_i - \mu_j \rangle}{\| \mu_i - \mu_j \|}
$$

Figures 8 and 9 show examples of quantum-like Gaussian mixture with small probability values between the two-dimensional Gaussians with $\mu_1 = (0, 1)^T$, $\Sigma_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ and $\mu_2 = (-1, -2)^T$, $\Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $\mu_3 = (2, -1)^T$, $\Sigma_3 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ and $\mu_4 = (-2, 2)^T$, $\Sigma_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with equal mixture coefficients $\pi$.

2.3 Switching Kalman filter

Kalman filter is an algorithm that uses a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables Russell and Norvig (2003). It estimates joint probability distribution over the variables for each timeframe.
Kalman filter makes a linear prediction. It assumes that the variables as well as the errors are random and Gaussian. A bird is heading at a high speed for a tree trunk, and the mean Gaussian as predicted by the Kalman filter will be centered on the obstacle, see Fig. 10. A bird would predict evasive action on one side or the other. A switching Kalman filter marginalizes the action of the bird; in our case, two different movements are described by the Gaussians. The marginalization is performed by the law of total probability resulting in the Gaussian mixture of the two Gaussians. The Gaussian mixture has a high probability values at the center of the obstacle. The quantum-like Gaussian mixture with the interference wave separates the two Gaussians, and the resulting probability value is low or non-present values at the center of the obstacle, see Figs. 10 and 11.

3 Gaussian mixture model

A Gaussian mixture model (GMM) is a probabilistic clustering model assumes that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters Bishop (2006). It incorporates information about the covariance structure of the data as well as the centers of the latent Gaussians. We will use the Gaussian mixture model for classification of the iris data set by comparing the obtained clusters with the actual classes from the data set. It contains 3 classes (Iris-Setosa, Iris-Versicolor, and Iris-Virginica ) of 50 instances each, where each class refers to a type of iris plant. Each instance is described by a four dimension vector the sepal and petal length and width Géron (2017). The use of this data set in cluster analysis is not common, since the data set only contains two clusters with rather obvious separation. One class is linearly separable from the other 2; the latter are not linearly separable from each other. As a result, two clusters are present. One of the clusters contains Iris-Setosa, while the other cluster contains both Iris-Versicolor and Iris-Virginica and is not separable. We will use the whole data set for learning and test since our task is to demonstrate the advantage of quantum-like mixture model for an unsupervised classifier. Between different Gaussians, interference is present that changes the shape of the quantum-like mixture distribution. A difference between the relative probabilities determined by the mixture Gaussians and the relative probabilities that are defined by quantum-like mixture Gaussians indicates the overlapping error that results from the striking distance to a decision boundary. For a Gaussian classifier, this area corresponds to the points that are close to the decision boundaries. Because of the resulting uncertainty, the points shouldn’t be assigned to any of the present classes.

3.1 Iris data set

In the first experiment we only take the first two features, sepal length and width, and plot the results in a two-dimensional coordinate system. We use the Scikit-learn Géron (2017) implementation to estimate Gaussian mixture model with full covariance matrices with following learned parameters
Quantum-like Gaussian mixture model

Fig. 8  a Four Gaussians. b Gaussian mixture of the four Gaussians with equal mixture coefficients $\pi$. c Quantum-like Gaussian mixture with the interference wave with $(\theta_i - \theta_j) = \frac{(\langle \mu_i | x | \mu_j \rangle - \langle \mu_j | x | \mu_i \rangle)}{||\mu_i - \mu_j||}$ of the four Gaussians with equal mixture coefficients $\pi$.

We use the model for classification of the iris data set by comparing the obtained clusters with the actual classes from the data set, see Fig. 12.

In Figs. 13 and 14 we indicate the resulting Gaussian mixture of the three Gaussians. One class is linearly separable from the other 2; the latter are not linearly separable from each other. As a result, two clusters are present with rather obvious separation. One of the clusters contains Iris-Setosa, while the other cluster contains both Iris-Versicolor and Iris-Virginica as indicated by the by the Gaussian mixture. The quantum-like Gaussian mixture separates the two clusters Iris-Versicolor and Iris-Virginica to some extent.

We use the model for classification of the iris data set by comparing the obtained clusters with the actual classes from the data set with the accuracy of 52.67%. The difference between the relative probability determined by the

\[
\pi_1 = 0.32444572, \mu_1 = \begin{pmatrix} 5.01507898 \\ 3.4514463 \end{pmatrix}, \\
\Sigma_1 = \begin{pmatrix} 0.11944714 & 0.08835648 \\ 0.08835648 & 0.11893388 \end{pmatrix}, \\
\pi_2 = 0.29474143, \mu_2 = \begin{pmatrix} 6.68055626 \\ 3.02849627 \end{pmatrix}, \\
\Sigma_2 = \begin{pmatrix} 0.36153508 & 0.05159664 \\ 0.05159664 & 0.08927917 \end{pmatrix},
\]

and

\[
\pi_3 = 0.38081285, \mu_3 = \begin{pmatrix} 5.9099976 \\ 2.74387546 \end{pmatrix}, \\
\Sigma_3 = \begin{pmatrix} 0.27671149 & 0.08897036 \\ 0.08897036 & 0.09398206 \end{pmatrix}.
\]
mixture Gaussian $\rho(x)$ and the relative probability $\rho_q(x)$ that is defined by quantum-like mixture Gaussian indicates the overlapping error that results from the striking distance of a decision boundary with

$$\rho(x) - \rho_q(x) \geq 0.006.$$ 

indicating that $x$ is too close to the decision boundary. Because of the resulting uncertainty, the point is not assigned to any of the present class represented by the Gaussian classifier. Through this constraint, 42 points were not assigned to any of the present class and the accuracy increased to 59.26%. With the constraint that the relative probability determined by the mixture Gaussian $\rho(x)$ should be above a certain threshold $\rho(x) > 0.04$, an optimum was reached for the threshold $\rho(x) > 0.04$, 8 points were not assigned to any of the present class, and the accuracy was only slightly improved 54.93%.

In the following experiment, we use all four features, the sepal and petal length and width. We use the Scikit-learn implementation to estimate Gaussian mixture model with full covariance matrices with following learned parameters

$$\pi_1 = 0.33333333, \quad \mu_1 = \begin{pmatrix} 5.006 \\ 3.428 \\ 1.462 \\ 0.246 \end{pmatrix},$$

$$\Sigma_1 = \begin{pmatrix} 0.121765 & 0.097232 & 0.016028 & 0.010124 \\ 0.097232 & 0.140817 & 0.011464 & 0.009112 \\ 0.016028 & 0.011464 & 0.029557 & 0.005948 \\ 0.010124 & 0.009112 & 0.005948 & 0.010885 \end{pmatrix}.$$
Quantum-like Gaussian mixture model

Fig. 11  a Two Gaussians with $\mu_1 = -2, \sigma_1 = 2$ and $\mu_2 = -2, \sigma_2 = 2$ representing the two different movements of the bird. b Gaussian mixture with equal mixture coefficients $\pi$ representing the switching Kalman filter. The resulting Gaussian mixture has a high probability values at the center corresponding to the obstacle. c Quantum-like Gaussian mixture with the interference wave with $\theta = \frac{(x-\mu_1)2\pi}{|\mu_1-\mu_2|}$ of the two Gaussians with $\mu_1 = -2, \sigma_1 = 1$ and $\mu_2 = -1, \sigma_2 = 1$ with equal mixture coefficients $\pi$. The resulting probability value of the center of the obstacle is low

$$\pi_2 = 0.30127092, \quad \mu_2 = \begin{pmatrix} 5.9170732 \\ 2.77804839 \\ 4.20540364 \\ 1.29848217 \end{pmatrix},$$

$$\Sigma_2 = \begin{pmatrix} 0.2755171 & 0.09662295 & 0.18547072 & 0.05478901 \\ 0.09662295 & 0.09255152 & 0.09103431 & 0.04299899 \\ 0.18547072 & 0.09103431 & 0.20235849 & 0.06171383 \\ 0.05478901 & 0.04299899 & 0.06171383 & 0.03233775 \end{pmatrix},$$

and

$$\rho_3 = 0.36539575, \quad \mu_3 = \begin{pmatrix} 6.54639415 \\ 2.94946365 \\ 5.48364578 \\ 1.98726565 \end{pmatrix},$$

$$\Sigma_3 = \begin{pmatrix} 0.38744093 & 0.09223276 & 0.30244302 & 0.06087397 \\ 0.09223276 & 0.11040914 & 0.08385112 & 0.05574334 \\ 0.30244302 & 0.08385112 & 0.32589574 & 0.07276776 \\ 0.06087397 & 0.05574334 & 0.07276776 & 0.08484505 \end{pmatrix}. $$

We use the model for classification of the iris data set by comparing the obtained clusters with the actual classes from the data set with the accuracy of 96.67%. The difference between the relative probability determined by the mixture Gaussian $\rho(x)$ and the relative probability $\rho_q(x)$ that is defined by

![Fig. 12 Classification of the first two features sepal length and width of the Iris data set by the Gaussian mixture model](image)
Fig. 13  a The three Gaussians. b Gaussian mixture of the three Gaussians. c Quantum-like Gaussian mixture with the interference wave with \( (\theta_i - \theta_j) = \frac{(|\mu_{i, j}|^2 - |\mu_i|^2)^{1/2} \cdot \pi}{|\mu_i - \mu_j|} \). The quantum-like Gaussian mixture separates the two clusters Iris-Versicolor and Iris-Virginica to some extent.

Fig. 14  a Contour plot of Gaussian mixture of three Gaussians. As a result, two clusters are present with rather obvious separation. One of the clusters contains Iris-Setosa, while the other cluster contains both Iris-Versicolor and Iris-Virginica as indicated by the Gaussian mixture. b Contour plot of quantum-like Gaussian mixture of the three Gaussians. The quantum-like Gaussian mixture separates the two clusters Iris-Versicolor and Iris-Virginica to some extent.
quantum-like mixture Gaussian indicates the overlapping error that results from the striking distance of a decision boundary with

\[
\rho(x) - \rho_q(x) \geq 0.002.
\]

indicating that the \( x \) is too close to the decision boundary. Because of the resulting uncertainty, the point is not assigned to any of the present class and the accuracy increased to 99.24%. With the constraint that the relative probability determined by the mixture Gaussian \( \rho(x) \) should be above a certain threshold, an optimum was reached for the threshold \( \rho(x) > 0.079 \), 26 points are not assigned to any of the present class, and the accuracy was only slightly improved 96.77%. The quantum-like mixture Gaussian can improve the classification accuracy by indicating that the uncertain points should not be assigned to any class.

4 Conclusion

A new concept of a quantum-like mixture model was introduced. It describes the mixture distribution with the assumption that a point is generated by each Gaussian at the same time. It was shown that the decision boundary of a quantum-like mixture Gaussian corresponds to the separation of probabilities for the switching Kalman filter. The quantum-like mixture Gaussian improves the classification accuracy in machine learning by indicating the uncertain points should not be assigned to any class. A difference between the relative probabilities determined by the mixture Gaussians and the relative probabilities that are defined by quantum-like mixture Gaussians indicates the overlapping error that results from the striking distance to a decision boundary. For a Gaussian classifier this area corresponds to the points that are close to the decision boundaries. Because of the resulting uncertainty, the points shouldn’t be assigned to any of the present classes. We demonstrated the principle by the quantum-like Gaussian mixture model for classification of the iris data set. The use of this data set in cluster analysis is not common, since the data set only contains two clusters with rather obvious separation. The accuracy of the mixture Gaussian model increased on the iris data set from 96.67% to 99.24%.

Acknowledgements We would also like to thank the anonymous reviewer for their valuable feedback. This work was supported by national funds through FCT, Fundação para a Ciência e a Tecnologia, Under Project UIDB/50021/2020.

Declarations

Conflict of interest The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors declare no conflicts of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

References

Binney J, Skinner D (2014) The physics of quantum mechanics. Oxford University Press, Oxford
Bishop C (2006) Pattern recognition and machine learning. Springer, Berlin
Busemeyer J, Wang Z (2014) Quantum cognition: key issues and discussion. Top Cogn Sci 6:43–46
Busemeyer JR, Wang Z, Townsend JT (2006) Quantum dynamics of human decision-making. J Math Psychol 50(3):220–241
Busemeyer JR, Wang Z, Lambert-Mogiliansky A (2009) Empirical comparison of Markov and quantum models of decision making. J Math Psychol 53(5):423–433
Busemeyer J, Pothos E, Franco R, Trueblood J (2011) A quantum theoretical explanation for probability judgment errors. J Psychol Rev 118:193–218
Busemeyer J, Wang Z, Trueblood J (2012) Hierarchical Bayesian estimation of quantum decision model parameters. In: Proceedings of the 6th International Symposium on Quantum Interactions. pp 80–89
Freier N (2013) Kalman filter: Multiplying normal distributions. Technical report, ITVS, TU Dresden
Géron A (2017) Hands-on machine learning with Scikit-learn and TensorFlow. O'Reilly, Newton
Khrennikov A (2009) Quantum-like model of cognitive decision making and information processing. J BioSyst 95:179–187
Russell SJ, Norvig P (2003) Artificial intelligence: a modern approach, 2nd edn. Prentice-Hall, Hoboken
Sergioli G (2020) Quantum and quantum-like machine learning: a note on differences and similarities. Soft Comput 24:10247–10255
Sergioli G, Militello C, Rundo L, Minafra L, Torrisi F, Russo G, Chow KL, Giuntini R (2021) A quantum-inspired classifier for clonal assay evaluations. Sci Rep 11:2830
Tversky A, Kahneman D (1973) Availability: a heuristic for judging frequency and probability. Cogn Psychol 5:207–232
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131
Wichert A (2020) Principles of quantum artificial intelligence: quantum problem solving and machine learning, 2nd edn. World Scientific, Singapore
Wichert A, Moreira C, Bruza P (2020) Quantum-like Bayesian networks. Entropy 22(2):170
Wichert A, Moreira C (2018) Balanced quantum-like model for decision making. In: Proceedings of 11th international conference on quantum interaction. Springer, editor, pp 79–90
Yukalov V, Sornette D (2011) Decision theory with prospect interference and entanglement. Theor Decis 70:283–328

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.