RESEARCH ARTICLE

Gaussian and Q-Gaussian Distributions Classification: a Heat Diffusion Kernel for Support Vector Machines

Luisa Toro Villegas[1*] | **Andrés Gómez Arango**[1*]

[1]Mathematics, EAFIT University, Medellin, Antioquia, 050022, Colombia

Correspondence
Email:

Funding information

This paper attempts to formulate a heat diffusion-based kernel for Support Vector Machines (SVMs). The paper explores the application of diffusion kernels on statistical manifolds, specifically focusing on the family of q-Gaussian distributions. The objectives include testing the diffusion kernel method in both univariate Gaussian and univariate q-Gaussian cases. Additionally, we compare the performance of diffusion kernels with linear, Gaussian, and Radial Basis Function (RBF) kernels in classification problems, specifically for classifying q-Gaussian generated data present in the $\Psi = (\mu, \sigma)$ variable space.

KEYWORDS
Diffusion Kernels, Classification, q-Gaussian Distributions, Support Vector Machines

1 | INTRODUCTION

Classification, a fundamental task in machine learning, involves assigning labels to objects based on their characteristics. Support Vector Machines, Cortes and Vapnik (1995), have emerged as powerful tools for solving classification problems. SVMs aim to find an optimal hyperplane that maximally separates data points belonging to different classes. This hyperplane is chosen based on the support vectors, which are the data points closest to the decision boundary.

In the context of SVMs, the kernel trick plays a crucial role, according to Amari and Wu (1999). Kernels enable SVMs to implicitly transform input data into a higher-dimensional space, allowing for the separation of non-linearly separable data. Common kernel functions include linear, polynomial, Gaussian (Radial Basis Function or RBF), and more. These kernels capture complex relationships within the data, enhancing the SVMs ability to generalize.

The methodology proposed builds upon these SVM principles, introducing a comprehensive approach to solving classification problems. The process involves separating the training set into subsets, choosing optimal hyper-planes or functions for non-separable cases, and incorporating soft separation and kernel functions for generalization.

Beyond traditional kernel functions, we delve into diffusion kernels applied to statistical manifolds. This exploration focuses on the family of q-Gaussian distributions, which provides a flexible framework for modeling probability distributions. Diffusion kernels on statistical manifolds leverage the geometric structure of the parameter space to enhance classification accuracy.

The objectives of this paper include testing the diffusion kernel method in univariate Gaussian and univariate q-Gaussian cases. Additionally, the aim is to compare the performance of diffusion kernels with linear, Gaussian, and RBF kernels in classification problems.

2 | METHODOLOGY

This document outlines the methodology for solving classification problems using kernel-based support vector machines (SVMs).

2.1 | Problem Formulation

Consider a classification problem with a function $F : X \to \{-1, 1\}$, where X is a set of objects. We have partial knowledge of F on a subset $T \subset X$, and the goal is to predict the remaining values of F.

2.2 | Abstract Representation

In practical terms, X often comprises non-numerical data, such as images. The function F assigns labels (-1 or 1) based on whether each datum possesses a specific property (e.g., an image showing a dog). To represent each datum, we define m attributes using real-valued functions $f_1, \ldots, f_m$ on X.

2.3 | Solution Approach

The methodology for solving the classification problem involves the following steps:

2.3.1 | Separating Training Set

If a separating hyperplane exists in $\mathbb{R}^m$ for subsets $T_+ = F^{-1}(1)$ and $T_- = F^{-1}(-1)$, an optimal hyperplane h_{opt} is chosen.

2.3.2 | Decision Function

For each $x \in X$, $F(x) = 1$ if x belongs to the half space determined by h_{opt} containing T_+, and $F(x) = -1$ otherwise.

If no separating hyperplane exists for T_+ and T_-, a function $\Phi : \mathbb{R}^m \to \mathbb{R}^k$ is chosen. Then, an optimal separating hyperplane h_{opt} is determined.

2.3.3 | Generalization

A generalization of the approach combines concepts of soft separation and kernel functions.

- *Soft Separation:* A penalty function P is introduced, assigning a real number to each hyperplane $h \subset \mathbb{R}^n$ indicating how well h separates two sets A and B. The lower the penalty $P(h)$, the better the separation.
- *Kernel Function:* A kernel function $\kappa : S \times S \to \mathbb{R}$ is introduced on a set S. The function κ is symmetric, and for any finite subset $R \subseteq S$, the matrix $(\kappa(x,y))_{x,y\in R}$ is positive semi-definite. The kernel function measures the similarity between elements in S. The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is defined as $k(x,y) = \langle \Phi(x), \Phi(y) \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product of $\mathbb{R}^k$. A function $k : \mathcal{X} \times \mathcal{X}$ is $k(x,y) = \langle \Psi(x), \Psi(y) \rangle$ for some $\Psi : \mathcal{X} \to \mathbb{R}^k$ if and only if k is a kernel function on $\mathcal{X}$.

2.4 | Methodology Steps

Given a classification problem $\mathcal{F} : X \to \{-1, 1\}$, a subset $\mathcal{T} = \{x(1), \ldots, x(N)\} \subset X$, and some knowledge of $\mathcal{F}|_{\mathcal{T}}$, the general method of solution involves the following steps:

- Choose a kernel function $\kappa : X \times X \to \mathbb{R}$.
- Consider the set D, such that

$$D = \left\{ P\left(\sum_{n=1}^{N} \beta_n \kappa(x, x(n)) + b \right) : \beta_1, \ldots, \beta_N, b \in \mathbb{R} \right\}$$

.

- Choose a penalization function $P : D \to \mathbb{R}$.
- Determine an element f of D minimizing P.
- Take $\hat{\mathcal{F}} = \operatorname{sgn}(f) : X \to \{-1, 1\}$ as the decision function, where $\operatorname{sgn}(x)$:

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

2.5 | Diffusion Kernels on Statistical Manifolds

The methodology proposed in Lafferty et al. (2005) introduces a natural kernel, termed the diffusion kernel, for classification problems. The classification problem is formulated within the framework of a parameterized family of probability distributions denoted as

$$P = \{ p(x, \xi) : \xi \in \Omega \subset \mathbb{R}^n \}$$

where Ω is an open set.

Suppose there are two independent sources $S1$ and $S2$ generating elements of X according to $p(\cdot, \xi_1)$ and $p(\cdot, \xi_2)$, respectively, where ξ_1, ξ_2 are known. Given a set of outcomes or data $D = \{x_1, \ldots, x_k\}$ originating from one of the sources (but it is unknown which one), the parameter ξ is estimated using appropriate statistics, yielding a point ξ_D in Ω. The task is to decide whether the data comes from $S1$ or $S2$ based on the value of ξ_D.

The proposed approach involves considering the Riemannian manifold of the smooth n-manifold P coordinatized by Ω, equipped with the Fisher information metric g. The heat kernel $K(\cdot, \cdot, t)$ for an appropriately chosen $t > 0$ is then employed as the kernel for the classification problem.

More concretely, the Riemannian manifold can be viewed as Ω endowed with the pullback g^* of the Fisher information metric. The heat equation $\partial u = \Delta g^* u / \partial t$ takes place in this Riemannian manifold. The heat kernel K is a real-valued map on $\Omega \times \Omega \times (0, \infty)$. For a training set $\{x_i\}$ with corresponding labels $y_i = f(x_i)$, each choice of $t > 0$ and nonnegative real values α_i determines a decision function $y_t : \Omega \to \mathbb{R}$, given by

$$y_t(x) = \sum_i \alpha_i y_i K_t(x, x_i).$$

Here, $K_t(\cdot, \cdot) = K(\cdot, \cdot, t)$. Note that $u(x,t) := y_t(x)$ is the solution to the heat equation on (Ω, g^*) with the initial condition $u(x,0) = \sum_i \alpha_i y_i \delta_{x_i}(x)$.

Although this paper primarily considers cases where the parameter space does not have a boundary, the ap-

proach presented in Lafferty et al. (2005) includes examples with manifolds having boundaries, such as multinomial families. The paper also discusses families of normal or Gaussian multivariate distributions.

2.6 | Objective

This paper aims to achieve two main objectives:

1. Test the diffusion kernel method in the univariate Gaussian case.
2. Test the diffusion kernel method in the univariate q-Gaussian case.

In the following sections, for each real $1 \leq q < 3$, a parameterized family denoted as N_q of probability distributions, known as q-Gaussian or q-normal univariate probability distributions, is introduced. The family of 1-Gaussian univariate probability distributions is equivalent to the family $N(1)$ of univariate Gaussian distributions.

The remainder of the paper focuses on comparing the performance of diffusion kernels with linear, Gaussian, and RBF kernels in classification problems.

2.7 | Family of q-Gaussian Distributions: Q-Deformed Characteristics

Let $q < 3$ be a real number. The q-Gaussian probability distribution with parameters $\xi = (\mu, \sigma)$, where $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$, is defined over the real line as

$$p(x, \xi) = \frac{1}{Z_{q,\sigma}} \exp_q \left(-\frac{(x-\mu)^2}{(3-q)\sigma^2} \right),$$

where $\exp_q$ is the q-exponential function and

$$Z = \int_{-\infty}^{\infty} \exp_q \left(-\frac{(x-\mu)^2}{(3-q)\sigma^2} \right) dx$$

Is the normalization constant.

And the q-exponential function is given by

$$\exp_q(x) = \begin{cases} \exp(x) & \text{if } q = 1, \\ (1 + (1-q)x)^{\frac{1}{1-q}} & \text{if } q \neq 1, \end{cases}$$

2.8 | Family of q-Gaussian Distributions

For each fixed real number $q < 3$, the family of q-Gaussian distributions is denoted as G_q and comprises all q-Gaussian probability distributions.

Note that the family G_1 of 1-Gaussian distributions encompasses both the family of usual Gaussian distributions and serves as the limit of G_q as $q \to 1$ in an appropriate sense. Specifically, it can be expressed as follows:

$$Z_{q,\sigma} = \begin{cases} [\sqrt{\frac{3-q}{1-q}} B(\frac{2-q}{1-q}, \frac{1}{2})]\sigma & \text{if } q < 1, \\ \sqrt{2\pi}\sigma & \text{if } q = 1, \\ [\sqrt{\frac{3-q}{q-1}} B(\frac{3-q}{2(q-1)}, \frac{1}{2})]\sigma & \text{if } 1 < q < 3. \end{cases} \quad (1)$$

Here, $Z_{q,\sigma}$ is the normalization constant, and B is the Beta function shown below:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$$

Similarly, just as Gaussian distributions maximize information entropy, for each $q < 3$, q-Gaussian probability distributions maximize Tsallis q-entropy.

2.9 | The Family of q-Gaussian Distributions as a Statistical Manifold

As mentioned in Section 2.5, every n-parametric family of probability distributions, where the parameter varies on an open subset Ω of $\mathbb{R}^n$, determines a Riemannian n-dimensional manifold. The points on this manifold are distributions coordinatized by the parameter and metrized by the classical Fisher information metric. In the case of G_q, the family of q-Gaussian distributions over the real line, there is an alternative Riemannian metric known as the q-Fisher information metric.

For each q-Gaussian family, except when $q = 1$, two distinct statistical 2-dimensional manifolds are determined. One is equipped with the classical Fisher information metric, denoted as g_q^F, and the other with the q-Fisher information metric, denoted as $g^{(q)}$. These two statistical manifolds can be viewed as the upper half plane of the μ, σ Cartesian plane, endowed with either of the metrics:

$$g_q^F = \frac{1}{q\sigma^2}(d\mu^2 + (3-q)d\sigma^2),$$
$$g^{(q)} = \frac{2}{(3-q)A_q^{1-q}\sigma^{3-q}}(d\mu^2 + (3-q)\,d\sigma^2).$$

For subsequent calculations, three metrics for the upper half μ, σ plane are considered:

$$g_{\text{euc}} = d\mu^2 + d\sigma^2,$$
$$g_{\text{euc},q} = \frac{1}{3-q}d\mu^2 + d\sigma^2,$$
$$g^H = \frac{1}{\sigma^2}(d\mu^2 + d\sigma^2).$$

In the following, two well-known facts about conformally equivalent metrics on a 2-dimensional manifold are utilized.

2.9.1 | Conformally Equivalent Metrics

Let g and h be two Riemannian metrics on the same 2-dimensional manifold, and assume they are conformally equivalent, i.e., $h = e^{2w}g$, where w is a smooth real-valued function on the manifold. Then, the Gaussian curvatures and the Laplacian-Beltrami operators are related by:

$$K(h) = e^{-2w}(K(g) - \Delta_g w),$$
$$\Delta_h = e^{-2w}\Delta_g.$$

Directly computing the Gaussian curvature of $g_{\text{euc},q}$ as $K(g_{\text{euc},q}) = 0$ and the Laplace-Beltrami operator induced by $g_{\text{euc},q}$ as $\Delta_{g_{\text{euc},q}} f = (3-q)f_{\mu\mu} + f_{\sigma\sigma}$, relations (9) and (10) imply:

$$K(g_q^F) = \frac{q\sigma^2}{3-q}(K(g_{\text{euc},q}) - \Delta_{g_{\text{euc},q}}(\frac{1}{2}ln(\frac{3-q}{q\sigma^2}))) = -\frac{q}{3-q} \tag{2}$$

and

$$\Delta_{g_q^F} f = \frac{q\sigma^2}{3-q}\Delta_{g_{euc,q}} f = \frac{q\sigma^2}{3-q}\left((3-q)f_{\mu\mu} + f_{\sigma\sigma}\right). \tag{3}$$

On the other hand, since $g^{(q)} = \frac{2q}{(3-q)A_q^{1-q}\sigma^{1-q}}\frac{3-q}{q\sigma^2}g_{euc,q} = \frac{2}{A_q^{1-q}\sigma^{3-q}}g_{euc,q}$, relations 9 and 10 say that

$$K\left(g^{(q)}\right) = \frac{A_q^{1-q}\sigma^{3-q}}{2}\left(K\left(g_{euc,q}\right) - \Delta_{g_{euc,q}}\left(\frac{1}{2}\ln\left(\frac{2}{A_q^{1-q}\sigma^{3-q}}\right)\right)\right) = -\frac{(3-q)A_q^{1-q}}{4\sigma^{q-1}} \tag{4}$$

and

$$\Delta_{g^{(q)}} f = \frac{A_q^{1-q}\sigma^{3-q}}{2}\Delta_{g_{euc,q}} f = \frac{A_q^{1-q}\sigma^{3-q}}{2}\left((3-q)f_{\mu\mu} + f_{\sigma\sigma}\right) \tag{5}$$

It is crucial to note the distinction between the Gaussian curvatures of g_q^F and $g^{(q)}$. While the Gaussian curvature of g_q^F is negative and constant, the curvature of $g_(q)$ is negative but non-constant, depending solely on σ.

We now aim to derive an explicit expression for the heat kernel $K_{g_q^F}$ on the upper half-plane μ, σ endowed with the metric g_q^F. Unfortunately, a similar expression for the heat kernel determined by the metric $g^{(q)}$ remains unknown to the authors.

To begin, consider the function $\phi(\mu, \sigma) = (\mu, \sqrt{3-q}\,\sigma)$, which acts as an isometry from the space (μ, σ) equipped with the metric g_q^F to the space (μ, σ) equipped with the metric $\frac{3-q}{q}g^H$. This implies that

$$K_{g_q^F}(x, x', t) = K_{\frac{3-q}{q}g^H}(\phi(x), \phi(x'), t). \tag{6}$$

Let us now calculate $K_{\frac{3-q}{q}g^H}(\phi(x), \phi(x'), t)$. To facilitate this calculation, we utilize the relation

$$K_{\frac{3-q}{q}g^H}(y, y', t) = K_{g^H}(y, y', \frac{q}{3-q}t), \tag{7}$$

whose validity we establish as follows.

Consider any metric g. If $u(x,t)$ is a solution to the equation $u_t = \Delta_g u$, if and only if $v(x,t) = u(x, \frac{t}{\alpha})$ is

a solution to the equation $v_t = \alpha^{-1}\Delta_g v = \Delta_{\alpha g} v$. This implies that if, at $t = 0$, there is a Dirac delta of heat at x_0, and we let it spread as if the geometry was g, and also as if the geometry was αg, then the spreading functions are related by $v(x,t) = u(x, \frac{t}{\alpha})$. If x_1 is another point, then $u(x_1, t) = u(x_1, \alpha(t/\alpha)) = v(x_1, \alpha t)$. We conclude that $K_{\alpha g}(x_0, x_1, t) = K_g(x_0, x_1, t/\alpha)$.

Combining these facts, we obtain the relation:

$$K_{g_q^{\sigma}}((\mu,\sigma),(\mu',\sigma'),t) = K_{g^H}(\phi(\mu),\phi(\sigma),\frac{q}{3-q}t) = K_{g^H}((\mu,\sqrt{3-q}\sigma),(\mu',\sqrt{3-q}\sigma'),\frac{q}{3-q}t). \quad (8)$$

Finally, it is worth noting that there exists an explicit formula for the heat kernel K_{g^H} (see [reference]):

$$K_{g^H}(y,y',t) = \frac{1}{2^{\frac{5}{2}}\pi^{\frac{3}{2}}t^{-\frac{3}{2}}} e^{-t/4} \int_{\rho}^{\infty} \frac{s \cdot e^{-\frac{s^2}{4t}}}{\sqrt{\cosh(s) - \cosh(\rho)}} ds \quad (9)$$

where $\rho = \text{dist}_{g^H}(y,y')$ represents the geodesic distance between y and y'.

3 | RESULTS AND ANALYSIS

Figure 1 is a scatter plot showing the spatial distribution of two sources generated for Experiment 1, where we know source 1 ($S1$ shown in green) has a mean $\mu = 0$, standard deviation $\sigma = 1$, and that comes from 2-Gaussian distribution. Source 2 ($S2$ shown in blue) has a mean $\mu = 2$, standard deviation $\sigma = 2$, and that comes from the same distribution.

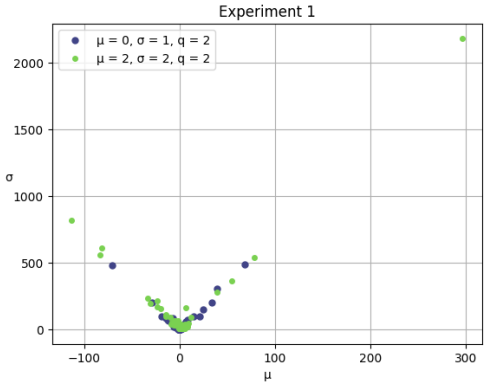

FIGURE 1 Experiment 1

The plot shows a strong overlap between the two sources. This represents a challenge for the classification method, due to the lack of separability, or at least at first glance.

From the points we get to observe, the only visual difference is that the V shape in $S2$ spans longer than $S1$.

Figure 2 shows the spatial distribution of two sources generated for Experiment 2, where $S1$ (shown in green) has a mean $\mu = 0$, standard deviation $\sigma = 1$, and that comes from 1-Gaussian distribution.$S2$ (shown in blue) has a mean $\mu = 10$, standard deviation $\sigma = 1$, and that comes from the same distribution.

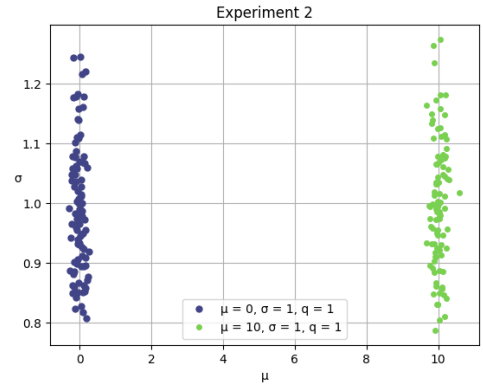

FIGURE 2 Experiment 2

The plot shows no overlap between the two sources. This is very easy to classify using any classification method, due to glaringly obvious separability.

From the points we get to observe, the space is neatly separated horizontally, with both sources spanning roughly the same amount of space and widely separated between them.

Figure 3 shows the spatial distribution of two sources generated for Experiment 3, where $S1$ (shown in green) has a mean $\mu = 0$, standard deviation $\sigma = 1$, and comes from 1.5-Gaussian distribution. $S2$ (shown in blue) has a mean $\mu = 0$, standard deviation $\sigma = 2$, and comes from the same distribution.

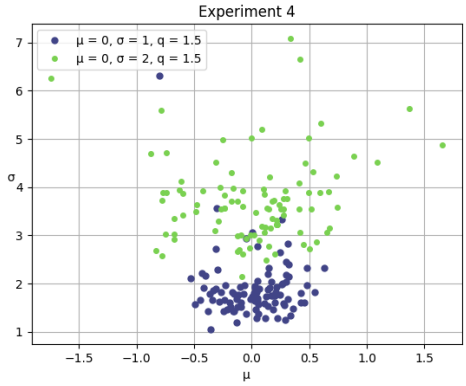

FIGURE 3 Experiment 3

The plot shows some overlap between the two sources. The separability between both is not as evident, and it may cause some misclassified points coming from the area of the overlap.

From the points we get to observe, the space is roughly separated vertically, but $S1$ occupies less space compared to $S2$. Tracing a separating hyper-plane for these points will prove a challenge.

Figure 4 shows where the SVM traced the hyper-plane if we were to visualize it in the original variable space for the best results obtained in the first image, for the worst results in the second one, and for the median results in the last image of the figure.

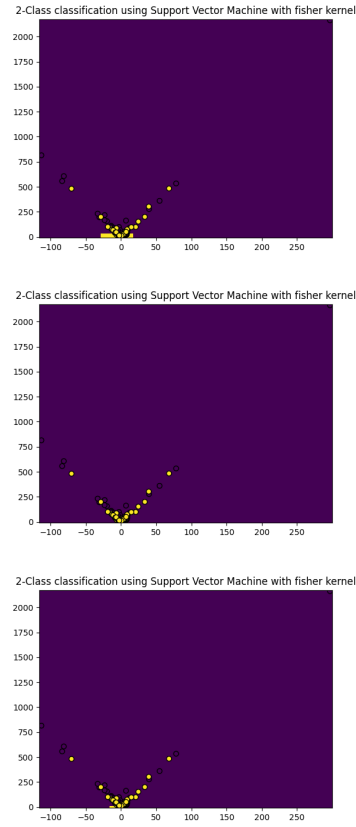

FIGURE 4 Experiment 1 SVM Classification with Fisher Kernels

Then in Figure 5 and Figure 6 it can be seen the same results yielded above but using the linear and RBF kernel respectively for experiment 1.

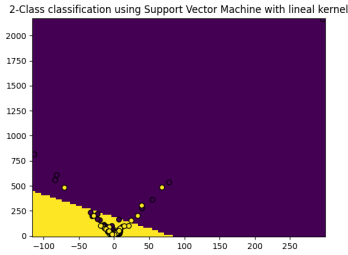

FIGURE 5 Experiment 1 SVM Classification with Linear Kernel

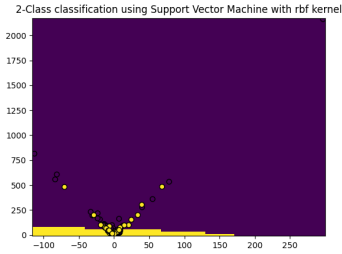

FIGURE 6 Experiment 1 SVM Classification with RBF Kernel

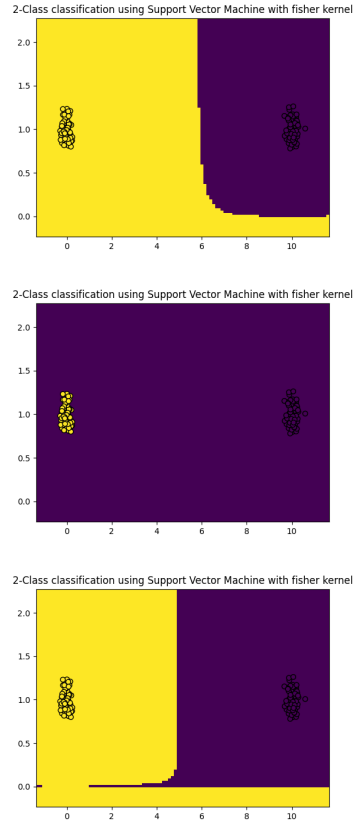

FIGURE 7 Experiment 2 SVM Classification with Fisher Kernels

Figure 7 shows where the SVM traced the hyperplane if we were to visualize it in the original variable space for the best results obtained in the first image, for the worst results in the second one, and for the median results in the last image of the figure.

Then in Figure 8 and Figure 9 it can be seen the same results yielded above but using the linear and RBF kernel respectively for experiment 2.

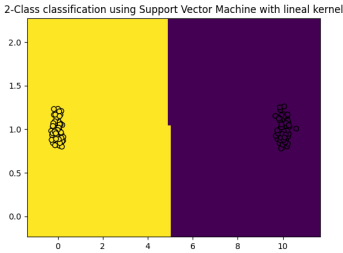

FIGURE 8 Experiment 2 SVM Classification with Linear Kernel

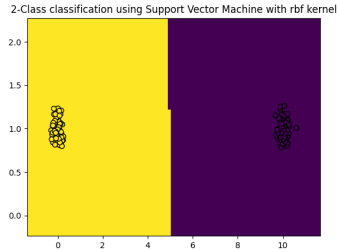

FIGURE 9 Experiment 2 SVM Classification with RBF Kernel

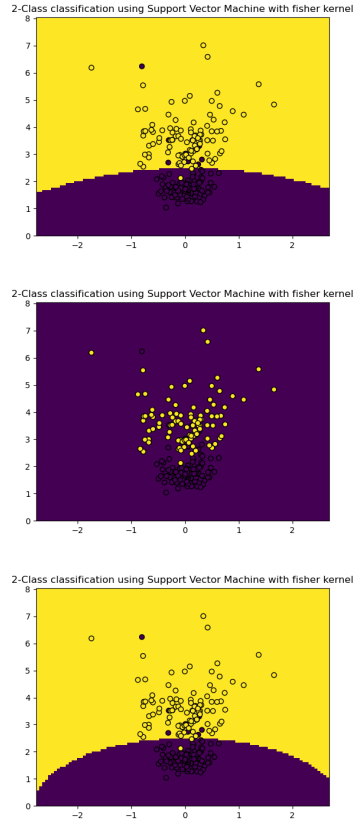

FIGURE 10 Experiment 3 SVM Classification with Fisher Kernels

Figure 10 shows where the SVM traced the hyperplane if we were to visualize it in the original variable space for the best results obtained in the first image, for the worst results in the second one, and for the median results in the last image of the figure.

Then in Figure 11 and Figure 12 it can be seen the same results yielded above but using the linear and RBF kernel respectively for experiment 3.

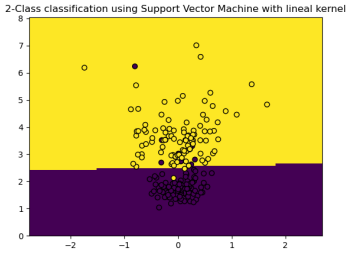

FIGURE 11 Experiment 3 SVM Classification with Linear Kernel

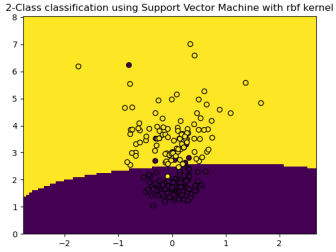

FIGURE 12 Experiment 3 SVM Classification with RBF Kernel

Overall, the Best Fisher method performed the best in all four experiments, followed by the Medium Fisher method. The Linear and RBF methods performed similarly, with the Linear method performing slightly better in Experiment 1. The Worst Fisher method performed poorly across all four experiments, as expected.

Figure 13 shows the results of SVM experiments for four experiments each having different sets of data. The authors attempted the following methods for classification: best fisher, worst fisher, median fisher, Linear, and RBF.

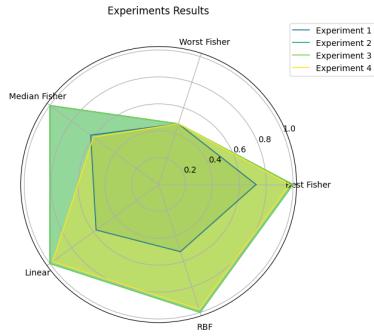

FIGURE 13 Experiment Results

Table 1 dives into the different hyper-parameters used to obtain the Fisher results in Figure 13 for each experiment.

Experiment	Chosen	t	q	Fisher
Experiment 1	Maximum	0.8	1	0.725
	Minimum	0.1	1	0.475
	Median	0.7	1.7	0.625
Experiment 2	Maximum	0.1	2.5	1
	Minimum	0.1	1	0.475
	Median	0.7	1.7	1
Experiment 3	Maximum	0.6	1.2	0.975
	Minimum	0.1	1	0.475
	Median	0.7	1.7	0.6

TABLE 1 Results shown on radar plot

4 | DISCUSSION AND CONCLUSION

In this paper, the efficacy of diffusion kernels was explored, specifically focusing on Gaussian and q-Gaussian distributions, within the realm of Support Vector Machines (SVMs). The primary objective was to rigorously test these diffusion kernels in both univariate Gaussian and q-Gaussian scenarios and compare their performance against traditional linear Gaussian and Radial Basis Function (RBF) kernels in classification tasks.

Our methodology leveraged the intrinsic geometric properties of statistical manifolds, particularly emphasizing the family of q-Gaussian distributions. This approach enabled us to utilize the geometric structure of the parameter space, thereby enhancing classification accuracy. The diffusion kernel's effectiveness was demonstrated through various experiments, where it showcased its superiority in scenarios where the parameter q was close to or equal to 1.

A notable observation from our experiments was that the diffusion kernels outperformed other kernels when q approached 1. However, as q increased, there was a discernible trend of diminishing performance for the diffusion kernel compared to linear and Gaussian kernels, especially in cases with larger values of σ^1 and σ^2. This suggests a nuanced behavior of diffusion kernels depending on the underlying distributional characteristics of the data.

The study's results have significant implications for the field of machine learning, particularly in classification tasks involving complex distributional data. The introduction of diffusion kernels based on the heat equation and their application in SVMs presents a novel way to enhance classification accuracy by capitalizing on the underlying statistical structure of the data.

In conclusion, our research successfully demonstrates the potential of diffusion kernels in SVM classification, especially in the context of q-Gaussian distributions. This work paves the way for further exploration into the application of advanced mathematical concepts like heat diffusion in machine learning, promising more refined and efficient classification models in future studies.

references

Amari, S.-i. and Wu, S. (1999) Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, **12**, 783–789.

Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Machine learning*, **20**, 273–297.

Lafferty, J., Lebanon, G. and Jaakkola, T. (2005) Diffusion kernels on statistical manifolds. *Journal of Machine Learning Research*, **6**.

A | RESULT TABLES

t	q	Fisher
0.1	1.0	0.475
0.1	1.2	0.475
0.1	1.5	0.475
0.1	1.7	0.475
0.1	2.0	0.475
0.1	2.5	0.475
0.1	2.8	0.475
0.5	1.0	0.625
0.5	1.2	0.65
0.5	1.5	0.625
0.5	1.7	0.525
0.5	2.0	0.525
0.5	2.5	0.525
0.5	2.8	0.65
0.6	1.0	0.625
0.6	1.2	0.675
0.6	1.5	0.625
0.6	1.7	0.625
0.6	2.0	0.625
0.6	2.5	0.625
0.6	2.8	0.675
0.7	1.0	0.7
0.7	1.2	0.625
0.7	1.5	0.675
0.7	1.7	0.625
0.7	2.0	0.625
0.7	2.5	0.625
0.7	2.8	0.65
0.8	1.0	0.725
0.8	1.2	0.625
0.8	1.5	0.625
0.8	1.7	0.65
0.8	2.0	0.65
0.8	2.5	0.675
0.8	2.8	0.675
0.9	1.0	0.65
0.9	1.2	0.675
0.9	1.5	0.6
0.9	1.7	0.625
0.9	2.0	0.625
0.9	2.5	0.65
0.9	2.8	0.7
1.0	1.0	0.575
1.0	1.2	0.725
1.0	1.5	0.675
1.0	1.7	0.625
1.0	2.0	0.6
1.0	2.5	0.625
1.0	2.8	0.675

TABLE 2 Results for experiment 1

t	q	Fisher
0.1	1.0	0.475
0.1	1.2	0.475
0.1	1.5	0.475
0.1	1.7	0.475
0.1	2.0	0.475
0.1	2.5	1.0
0.1	2.8	1.0
0.5	1.0	1.0
0.5	1.2	1.0
0.5	1.5	1.0
0.5	1.7	1.0
0.5	2.0	1.0
0.5	2.5	1.0
0.5	2.8	1.0
0.6	1.0	1.0
0.6	1.2	1.0
0.6	1.5	1.0
0.6	1.7	1.0
0.6	2.0	1.0
0.6	2.5	1.0
0.6	2.8	1.0
0.7	1.0	1.0
0.7	1.2	1.0
0.7	1.5	1.0
0.7	1.7	1.0
0.7	2.0	1.0
0.7	2.5	1.0
0.7	2.8	1.0
0.8	1.0	1.0
0.8	1.2	1.0
0.8	1.5	1.0
0.8	1.7	1.0
0.8	2.0	1.0
0.8	2.5	1.0
0.8	2.8	1.0
0.9	1.0	1.0
0.9	1.2	1.0
0.9	1.5	1.0
0.9	1.7	1.0
0.9	2.0	1.0
0.9	2.5	1.0
0.9	2.8	1.0
1.0	1.0	1.0
1.0	1.2	1.0
1.0	1.5	1.0
1.0	1.7	1.0
1.0	2.0	1.0
1.0	2.5	1.0
1.0	2.8	1.0

TABLE 3 Results for experiment 2

t	q	Fisher
0.1	1.0	0.475
0.1	1.2	0.475
0.1	1.5	0.475
0.1	1.7	0.475
0.1	2.0	0.475
0.1	2.5	0.475
0.1	2.8	0.475
0.5	1.0	0.925
0.5	1.2	0.925
0.5	1.5	0.6
0.5	1.7	0.5
0.5	2.0	0.5
0.5	2.5	0.5
0.5	2.8	0.675
0.6	1.0	0.775
0.6	1.2	0.975
0.6	1.5	0.825
0.6	1.7	0.575
0.6	2.0	0.525
0.6	2.5	0.525
0.6	2.8	0.925
0.7	1.0	0.625
0.7	1.2	0.9
0.7	1.5	0.95
0.7	1.7	0.825
0.7	2.0	0.6
0.7	2.5	0.65
0.7	2.8	0.95
0.8	1.0	0.975
0.8	1.2	0.8
0.8	1.5	0.5
0.8	1.7	0.5
0.8	2.0	0.475
0.8	2.5	0.475
0.8	2.8	0.525
0.9	1.0	0.925
0.9	1.2	0.925
0.9	1.5	0.6
0.9	1.7	0.5
0.9	2.0	0.5
0.9	2.5	0.5
0.9	2.8	0.6
1.0	1.0	0.775
1.0	1.2	0.975
1.0	1.5	0.825
1.0	1.7	0.575
1.0	2.0	0.525
1.0	2.5	0.525
1.0	2.8	0.925

TABLE 4 Results for experiment 3