Manifold-Adaptive Metric Learning via Locally Consistent Weighting

Yuying Huang
Maynooth International Engineering College, Fuzhou, China

Abstract

Accurate multi-class classification in complex, multidimensional sensory datasets, such as the assessment of wine quality based on physicochemical properties, presents a significant challenge for data-driven modelling. The core difficulty lies in predicting quality grades from multiple chemical attributes, a task characterised by inherent class imbalance—higher quality grades are often underrepresented—and substantial overlap in the feature distributions between adjacent quality categories. In response to these difficulties, we introduce Manifold-Corrected Adaptive Weighted KNN(MCAW-KNN), an adaptive weighting scheme based on local contextual learning. The central premise of MCAW-KNN is the dynamic identification of a high-quality local neighbourhood surrounding a test instance. This neighbourhood is constructed to be representative of the underlying data manifold while maintaining high proximity to the instance in question.

Keywords: manifold-adaptive; weighting scheme; metric learning

Introduction

Conventional feature-weighting approaches, which derive global weights from models like Support Vector Machines (SVMs)[1] or Random Forests[2], are fundamentally inadequate. They assume a uniform feature importance across the entire operational space, an assumption that fails under the localised data structures and imbalances intrinsic to fault diagnostics, leading to suboptimal classification performance[3-10].

The concept of local linear modelling demonstrates strong potential by accurately characterising the linear or piecewise linear relationships of local data structures and generalising them to global decisions. The effectiveness of local modelling for global generalisation is the key to tackle the challenge of data complexity and class imbalance. As inspired by the success of local linear embedding (LLE)[11-13] in preserving global topological structure, the proposed Manifold-Corrected Adaptive Weighted KNN (MCAW-KNN) algorithm is developed for the question with these limitations. The central premise is the dynamic identification of a high-quality, representative local neighbourhood for each test instance. This neighbourhood is not merely defined by simple proximity but is strategically constructed to reflect the underlying data manifold while ensuring relevance to the test point. Two methods are introduced for this purpose: one leveraging a genetic algorithm for optimal neighbourhood search, and another employing a statistical model based on reverse neighbour rank and class compactness. Within this optimised local context, a discriminative, instance-specific feature weight matrix is derived with global manifold correction, to compute the distance metric with the local geometry and global generalisation.

Manifold-Corrected Adaptive Weighted KNN(MCAW-KNN)

Discriminant Adaptive Neighbourhood

The MCAW-KNN algorithm departs from the conventional passive selection of a test point's nearest neighbours. Instead, it actively employs an algorithm to search for and construct a high-quality local neighbourhood around the test point, as a local context to better reflect the identity of test case. Two search strategies are applied for discussion in this work: search by genetic algorithm and search by bidirected neighbour saliency.

Search by Genetic Algorithm: The genetic algorithm's ability to converge towards a near-optimal solution is grounded in the Schema Theorem and the Building Block Hypothesis[14][15]. These theorems posit that genetic algorithms efficiently combine short, high-fitness gene segments ("building blocks") to progressively construct solutions approximating the Pareto front of a multi-objective optimisation problem. The evaluation function in MCAW-KNN incorporates two key objectives: distance similarity and class purity.

Distance Similarity: The average distance between each individual in the population (i.e., a candidate set of neighbourhood samples) and the test point should be minimised. This ensures the local relevance of the neighbourhood.

Class Distribution Purity: The consistency between the class distribution within an individual and the global dataset's class distribution is incorporated into the evaluation. This helps to select sample regions with clear, unambiguous class labels, thereby enhancing the quality of the neighbourhood rather than merely pursuing proximity. The use of evolutionary algorithms to optimise KNN parameters, such as feature weights or instance selection, has been demonstrated to effectively enhance classification performance [16-22], particularly on imbalanced datasets. This work extends the concept to the active construction of the neighbourhood itself.

Search by Bidirected Neighbour Saliency: In the process of identifying neighbours for a test instance, the algorithm ex-

tends beyond relying solely on the conventional forward geometric distance—the traditional measure from a sample to the test point. It critically incorporates an assessment of its local representativeness within its own class environment through a reverse nearest neighbour ranking. Concurrently, the internal distribution characteristics of the sample's class are quantified by computing its class compactness. These three elements—forward distance, reverse ranking, and class compactness—are integrated within a cascaded model. This model employs Bayesian smoothing[23-25] to address uncertainties inherent in small-sample scenarios and utilizes Wilson score interval estimation[26][27] to provide a conservative assessment. The synthesis culminates in the generation of a smooth adjustment factor specific to each candidate sample. This factor dynamically recalibrates the original distance metric, thereby favouring the selection of samples that exhibit both high representativeness within their respective class and close proximity to the test point, and these samples ultimately contribute to the formation of a high-quality local neighbourhood.

Global Manifold Constraint for Correcting Local Geometry

To mitigate potential sampling bias or geometric distortion in the local neighbourhood, MCAW-KNN introduces a global manifold[28] constraint to correct the weight matrix calculated based on the local neighbourhood. This constraint is derived from the generalised Rayleigh quotient computed over the entire dataset, which captures the feature directions most discriminative for class separation at a global scale[29][30]. This process can be conceptualised as "projecting" the local neighbourhood onto a more stable subspace guided by the global discriminative manifold. This step ensures that decisions based on the local neighbourhood remain consistent with the global distribution structure of the data, significantly enhancing the algorithm's robustness against noise and outliers.

Rayleigh Quotient for Class-Aware Weights

The generalised Rayleigh quotient[30] serves as the core mathematical tool for obtaining the optimal weight vector. For the local region identified by the genetic algorithm, the algorithm performs a one-versus-rest binary classification task for each class. It constructs the between-class and within-class scatter matrices specific to that local region. The eigenvector corresponding to the maximisation of the generalised Rayleigh quotient represents the feature importance metric that best discriminates the current class from all other classes within that local context. This allows the model to adaptively emphasise the features most critical for distinguishing a particular class.

Class-Based Weight Matrix and Recall Score Integration

A fundamental limitation of the traditional KNN algorithm lies in its use of a single, unified distance metric (e.g., Mahalanobis distance) for all classes. This implicitly assumes that all features contribute equally to distinguishing any class, which often does not hold in practice, as the key discriminatory attributes can vary significantly between classes. To address this, MCAW-KNN adopts a class-specific KNN approach and a dual-weighted voting mechanism that considers both sample-level similarity and class-level representativeness[31]. The voting weight of a neighbour for a given test sample is determined by two components:

Class-Sensitive Distance Weight: The distance is computed using the unique weight matrix of the class to which the neighbour belongs. This weighting ensures that the vote relies more heavily on the features most effective for distinguishing that neighbour's class.

Class Recall Score Weight: The frequency (i.e., recall rate) at which the neighbour's class appears within the high-quality neighbourhood identified by the genetic algorithm serves as a measure of that class's overall support and reliability within the local context.

Theoretical Properties

In this section, we analyse the mathematical concepts and theoretical feasibility of MCAW-KNN. Let the test instance be x_0 , and let the local neighbourhood $\mathcal{N}(\mathbf{x}_0)$ be the set of selected data points. We characterise this region by its empirical covariance matrix Σ_N.

Geometric Interpretation of Tangent Space

The selected local region $\mathcal{N}(\mathbf{x}_0)$ can be geometrically interpreted as a discrete approximation of the tangent space to the data manifold at the point $\mathbf{x}_0$. On a smooth, continuous data manifold $\mathcal{M}$, the tangent space $T_{\mathbf{x}_0}\mathcal{M}$ provides the best linear approximation of $\mathcal{M}$ within an infinitesimal neighbourhood of $\mathbf{x}_0$. Any vector $\mathbf{v} \in T_{\mathbf{x}_0}\mathcal{M}$ represents an instantaneous direction of change. A key property is that the inner products—and consequently the relative distances—of points projected onto this tangent space are preserved under manifold motions. In our discrete data scenario, we construct this "tangent space" by performing principal component analysis (PCA) on the points within $\mathcal{N}(\mathbf{x}_0)$. The principal components (the eigenvectors of the empirical covariance matrix Σ_N) form an orthonormal basis that spans the dominant linear subspace of the local region. By imposing the idealised requirement that $\Sigma_{\mathcal{N}} \approx \Sigma_{\mathcal{G}}$ (where $\Sigma_{\mathcal{G}}$ is a global reference covariance), we align this local linear subspace with the overall geometric structure of the data manifold. The resulting subspace, from $\mathcal{N}(\mathbf{x}_0)$, thus functions as a faithful discrete tangent space. Consequently, when data points near $\mathbf{x}_0$ are projected onto this local PCA space, their relative distances accurately reflect the true geodesic distances along the underlying manifold. This provides a rigorous geometric foundation for performing subsequent linear discriminant analysis within this space.

Class-Specific Weights from the Generalised Rayleigh Quotient

Within this localised and geometrically faithful space $\mathcal{N}(\mathbf{x}_0)$, we learn an independent attribute weight vector for each class C_i. This is achieved by formulating a binary classification problem: instances belonging to C_i versus those not belonging to C_i (denoted $-C_i$). The optimal weight vector $\mathbf{w}_i$ for class C_i is obtained by maximising the Generalised Rayleigh Quotient.

Let $S_b^{(i)}$ and $S_w^{(i)}$ be the between-class and within-class scatter matrices, respectively, computed for the C_i vs. $\neg C_i$ dichotomy within $\mathcal{N}(\mathbf{x}_0)$. We maximise the objective function $J(\mathbf{w}_i)$

$$J(\mathbf{w}_i) = \frac{\mathbf{w}_i^T S_b^{(i)} \mathbf{w}_i}{\mathbf{w}_i^T S_w^{(i)} \mathbf{w}_i}$$

Maximising $J(\mathbf{w}_i)$ yields an optimal projection direction $\mathbf{w}_i^*$ that simultaneously maximises the separation between class C_i and $-C_i$ while minimising the variance within each group. The solution to this optimisation problem is given by the generalised eigenvalue equation:

$$\mathbf{S}_i^{(j)} \mathbf{w}_i = \lambda \mathbf{S}_w^{(j)} \mathbf{w}_i$$

The eigenvector corresponding to the largest eigenvalue λ^* is the optimal weight vector $\mathbf{w}_i^*$. Crucially, this vector defines a new distance metric within the local tangent space. The distance from the test point $\mathbf{x}_0$ to any other point $\mathbf{x}$ is then computed as $|\mathbf{w}_i^* T(\mathbf{x} - \mathbf{x}_0)|$, representing the distance in the one-dimensional subspace that provides optimal discriminative power for class C_i.

Class Weight as Discriminant Basis Vectors

Following the preceding construction, we perform an eigendecomposition of the local covariance matrix $\Sigma_{\mathcal{N}} = \mathbf{V}\Lambda\mathbf{V}^T$. This yields a set of orthonormal basis vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_l\}$, which correspond to the directions of maximum data variance within $\mathcal{N}(\mathbf{x}_0)$. These PCA basis vectors describe the intrinsic data distribution. For the specific task of classification, however, we do not directly use these variance-maximizing directions. Instead, for each class C_i, we learn a specific weight vector $\mathbf{w}_i$. While the optimal direction $\mathbf{w}_i^*$ may not coincide with any single original PCA eigenvector, it is constrained to lie within the same tangent space. Therefore, it can be expressed as a linear combination of the primary PCA basis vectors:

$$\mathbf{w}_i = \sum_{j=1}^{d} \alpha_j \mathbf{v}_j$$

The process of solving the generalized eigenvalue problem effectively determines the combination coefficients α_i with the explicit objective of maximising class separability in this localised context. Consequently, each learned weight vector $\mathbf{w}_i$ can be interpreted as a custom-designed discriminant basis vector, tailored specifically for distinguishing class C_i within this local subspace, and normalised such that $\|\mathbf{w}_i\|_c = 1$. This vector captures the predominant patterns of data variation within $\mathcal{N}(\mathbf{x}_0)$ that are most effective for differentiating instances of C_i from all other classes.

The weight vector associated with an attribute for a class C_i is defined as $\mathbf{w}_c = (w_{c1}, w_{c2}, ..., w_{cn})^T$. This vector essentially scales the original feature space according to the learned discriminative properties of class C_i. If a test point $\mathbf{x}_{test}$ and a training point $\mathbf{x}_{train}^{(c)}$ both belong to C_i, their feature values along each original dimension i are scaled by a factor of $\sqrt{w_{ci}}$. Distances between such scaled points are measured within the discriminant subspace defined by:

$$d_c\left(\mathbf{x}_{test}, \mathbf{x}_{train}^{(e)}\right) = \sqrt{\sum_{i=1}^{n} w_{ci} \cdot \left(x_{test,i} - x_{train,i}^{(e)}\right)^2}$$

When a training sample $\mathbf{x}_{train}^{(-e)}$ does not belong to the target class C_i, it is still necessary to evaluate its distance from the test point $\mathbf{x}_{test}$ under the metric specific to class C_i. This is because the classification decision for the test point should be made by assessing all its neighbours using a consistent, class-discriminative metric. Within this algorithm, the weighted distance is essentially the absolute difference between the projections of the two points onto the discriminant basis vector $\mathbf{w}_c$. This projection distance reflects their separation along the direction defined by $\mathbf{w}_c$, which is optimised to maximise the margin between classes. Ideally, under this optimal projection, samples from the same class will have tightly clustered projections, while those from different classes will be far apart. The projection distance is defined as:

$$d_c(\mathbf{x}_{test}, \mathbf{x}_{train}) = |\mathbf{w}_c^T \mathbf{x}_{test} - \mathbf{w}_c^T \mathbf{x}_{train}|$$

Theorem 1. In the projection space defined by C_i, the expected distance between a test point $\mathbf{x}_{test}$ and a training point from a different class $\mathbf{x}_{train}^{(\neg c)}$ is, on average, greater than the expected distance to that from the same class $\mathbf{x}_{train}^{(c)}$

Proof. Let the test point $\mathbf{x}_{test}$ belong to class C_i. For a point $\mathbf{x}_{train} \in c$, the expected projection distance approximates the internal variance of the class projections along w. Crucially, one of the design objectives for w_c is to minimise the within-class scatter $\mathbf{w}_c^T \mathbf{S}_W \mathbf{w}_c$, which directly implies a small variance for the projections of samples from the same class. Consequently, $\mathbb{E}[d_c(\mathbf{x}_{test}, \mathbf{x}_{train} \in c)]$ is small. For a point $\mathbf{x}_t rain \notin c$, the class mean $\mathbf{m}_{-c}$ differs from the test point's class mean $\mathbf{m}_c$. The complementary objective for w is to maximise the between-class scatter $\mathbf{w}_c^T \mathbf{S}_B \mathbf{w}_c$, which is equivalent to maximising $|\mathbf{w}_c^T(\mathbf{m}_c - \mathbf{m}_{-c})|$. This induces a significant systematic offset between the projection of the test point and the projections of points from other classes. Therefore, $\mathbb{E}[d_c(\mathbf{x}_{test}, \mathbf{x}_{train} \notin c)]$ is comparatively large. Since the optimal weight vector $\mathbf{w}_c^*$ is derived precisely to achieve the dual aim of minimising within-class scatter and maximising between-class scatter, the inequality stated in the theorem holds at the optimum:“

$$\mathbb{E}[d_c(\mathbf{x}_{test}, \mathbf{x}_{train} \in c)] < \mathbb{E}[d_c(\mathbf{x}_{test}, \mathbf{x}_{train} \notin c)]$$

For a test point $\mathbf{x}_{test}$ not belonging to class $C_i, \mathbf{x}_{test}$ can be orthogonally decomposed inta the direction of $\mathbf{w}_c$ and the direction perpendicular to $\mathbf{w}_c$,and thus the inequality of the projection distance can be obtained. ‘

Manifold Constraints and Correction

In an ideal scenario where the data forms a continuously varying convex set, if one defines a local neighbourhood $\mathcal{N}(\mathbf{x}_0)$ such that its covariance matrix equals the global covariance matrix, i.e., $\Sigma_{\mathcal{N}} = \Sigma_{\mathcal{G}}$, the local "tangent space" would perfectly capture the global geometry, and relative distances between any two points projected onto this space would be preserved. Such a system could be considered "ideal". However, real-world data is discrete, and its partitioning inherently involves subjective variance. Systematic errors arise from the difficulty in finding a local neighbourhood that satisfies $\Sigma_{\mathcal{N}} = \Sigma_{\mathcal{G}}$. This error has two primary sources: firstly, the covariance matrix $\Sigma_{\mathcal{N}}$ of an empirically obtained local neighbourhood generally does not equal $\Sigma_{\mathcal{G}}$; secondly, while a covariance matrix fitted to a local region around test point $\mathbf{x}_0$ may yield features with good global generalisability, transient data fluctuations can still cause variability within an infinitesimal vicinity of the test point. Therefore, our approach is not to seek the "ideal tangent space" at $\mathbf{x}_0$ directly, but rather to find an "approximate tangent space" and then refine it by aligning it with global manifold constraints. Consider a general case: a global dataset is defined by $\mathcal{G} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^d$ and the labels y_i belong to a set of classes $C = \{C_1, C_2, ..., C_K\}$. The goal is to find a high-quality local neighbourhood $\mathcal{N}^*(\mathbf{x}_0)$ that satisfies a multi-objective evaluation criterion $F(\mathcal{N}, \mathbf{x}_0)$. Drawing inspiration from the Nacozy manifold correction concept from celestial mechanics, the local "approximate tangent space" at the test point $\mathbf{x}_0$ is fitted to the global manifold constraints.

$$\mathcal{N}^*(\mathbf{x}_0) = \arg\max_{\mathcal{N} \subseteq \mathcal{G}} F(\mathcal{N}, \mathbf{x}_0)$$

$$\mathbf{w}_{\text{corrected}}^{(k)} = \Phi_{\text{manifold}}(\mathbf{w}_{\text{local}}^{(k)}, \Sigma_{\mathcal{N}}, \Sigma_{\mathcal{G}})$$

Theorem 2. (Ideal Tangent Space Approximation) If the optimal local neighbourhood satisfies $\Sigma_{\mathcal{N}^*} = \Sigma_{\mathcal{G}}$, then the corrected weight vector equals the locally learned weight vector, i.e, $\mathbf{w}_{\text{corrected}}^{(k)} = \mathbf{w}_{\text{local}}^{(k)}$. More generally, the corrected weight vector $\mathbf{w}_{\text{corrected}}^{(k)}$ approximates the gradient of the class C_k discriminant function on the global manifold with respect to the metric defined by the covariance structure:

$$\mathbf{w}_{\text{corrected}}^{(k)} \approx \nabla_{\Sigma_a} f^{(k)}(\Sigma)$$

where $f^{(k)}(\Sigma)$ is the discriminant function for class C_k under the metric induced by covariance matrix Σ.

The manifold correction is intrinsically linked to the Generalised Rayleigh Quotient. The within-class scatter matrix $\mathbf{S}_w$ is fundamentally a weighted sum of the per-class covariance matrices. Therefore, correction based on the Rayleigh Quotient manifold inherently performs covariance-based correction. Furthermore, by incorporating the between-class scatter matrix $\mathbf{S}_b$, the global Rayleigh Quotient manifold provides a tighter and more discriminative constraint for per-class.

The process of correcting the locally learned weight vector $\mathbf{w}_{\text{local}}$ onto the target manifold defined by the global data's Generalised Rayleigh Quotient can be formulated as a constrained minimisation problem:

$$\text{Minimize}: \quad \frac{1}{2}(\mathbf{w} - \mathbf{w}_{local})^T (\mathbf{w} - \mathbf{w}_{local})$$
$$\text{Subject to}: \quad \Phi(\mathbf{w}) = 0$$

Here, the objective is to find the smallest perturbation to the local weight vector such that the resulting vector $\mathbf{w}_{corrected}$ lies on the global discriminant manifold, enforced by the constraint $\Phi(\mathbf{w}) = 0$.

By introducing Lagrange multipliers, specifically, for a class C_i , an "ideal" weight vector $\mathbf{w}_c^*$ should satisfy the generalised eigenvalue equation formed by the global scatter matrices: $\mathbf{S}_b^{global}\mathbf{w}_c^* = \lambda \mathbf{S}_w^{global}\mathbf{w}_c^*$. Here, $\mathbf{S}_b^{global}$ and $\mathbf{S}_w^{global}$ are the between-class and within-class scatter matrices computed from the global training data. The objective of this process is to find a new vector $\mathbf{w}_c^{corrected}$ that is as close as possible to the original local weight vector $\mathbf{w}_c^{local}$ while simultaneously satisfying or approximately satisfying the global manifold constraints The mathematical formulation for this correction is given by:

$$\mathbf{w}_c^{corrected} = \arg\ \min_{\mathbf{w}} \tfrac{1}{2}||\mathbf{w} - \mathbf{w}_c^{local}||^2 + \lambda \cdot ||\mathbf{S}_b^{qlobal}\mathbf{w} - \rho(\mathbf{w})\mathbf{S}_w^{global}\mathbf{w}||^2$$

Theorem 3 (Performance Lower Bound).

Assume the corrected weight vector $\mathbf{w}_c^{corrected}$ satisfies the manifold constraint $\|\mathbf{S}_b^{global}\mathbf{w} - \rho(\mathbf{w})\mathbf{S}_w^{global}\mathbf{w}\|^2 \leqslant \epsilon$. Then, its generalised Rayleigh quotient on the global data $J_{global}(\mathbf{w}_c^{\text{corrected}})$, satisfies the following relationship:

$$J_{global}(\mathbf{w}_c^{\text{corrected}}) = \frac{(\mathbf{w}_c^{\text{corrected}})^T \mathbf{S}_b^{\text{global}} \mathbf{w}_c^{\text{corrected}}}{(\mathbf{w}_c^{\text{corrected}})^T \mathbf{S}_w^{\text{global}} \mathbf{w}_c^{\text{corrected}}} \geqslant \rho - \delta(\epsilon)$$

where $\delta(\epsilon)$ is a quantity that approaches zero as ϵ decreases.

Proof. This lower bound can be derived by analysing the intrinsic meaning of the constraint. Here, ρ is reasonably set to $\rho = J_{global}(\mathbf{w}_c^{local})$, which is the value of the generalised Rayleigh quotient computed on the global data for the original local weight vector $\mathbf{w}_c^{local}$. The significance of this choice is that it establishes a comparative baseline. This implies that provided the correction process is sufficiently accurate (i.e., ϵ is small enough), the global discriminative performance of the corrected vector is guaranteed to be no worse than that of the original local vector, ensuring the "safety" of the correction operation.‘ Let us consider the general case of a soft constraint, where the constraint condition is: $\mathbf{w}^T\mathbf{S}_b^{global}\mathbf{w} - \rho\mathbf{w}^T\mathbf{S}_w^{global}\mathbf{w} \geqslant -\epsilon$, with $\epsilon > 0$ being a small constant representing the permitted degree of constraint violation. The optimisation problem thus becomes:

$$\mathbf{w}_c^{corrected} = \arg\ \min_{\mathbf{w}} \frac{1}{2}||\mathbf{w} - \mathbf{w}_c^{local}||^2$$

Subject to $\mathbf{w}^T\mathbf{S}_b^{\text{global}}\mathbf{w} - \rho\mathbf{w}^T\mathbf{S}_w^{\text{global}}\mathbf{w} \geqslant -\epsilon$ Construct the Lagrangian function, with $\lambda > 0$:

$$\mathcal{L}(\mathbf{w}, \lambda) =$$
$$\frac{1}{2}||\mathbf{w} - \mathbf{w}_c^{\text{local}}||^2 - \lambda \cdot \left(\mathbf{w}^T\mathbf{S}_b^{\text{global}}\mathbf{w} - \rho\mathbf{w}^T\mathbf{S}_w^{\text{global}}\mathbf{w} + \epsilon\right)$$

The Lagrangian $\mathcal{L}(\mathbf{w}, \lambda)$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions at the optimal solution $\left(\mathbf{w}_c^{\text{corrected}}, \lambda^*\right)$. When $\lambda^* = 0$, we obtain $\mathbf{w}_c^{\text{local}} = \mathbf{w}_c^{\text{corrected}}$, leading to $J_{\text{global}}\left(\mathbf{w}_c^{\text{corrected}}\right) = J_{\text{global}}\left(\mathbf{w}_c^{\text{local}}\right) = \rho$. It is then evident that $J_{\text{global}}\left(\mathbf{w}_c^{\text{corrected}}\right) \geqslant \rho - \delta(\epsilon)$. We now examine the case where $\lambda^* > 0$. Let $B = \mathbf{w}^T\mathbf{S}_b^{\text{global}}\mathbf{w}$ and $W = \mathbf{w}^T\mathbf{S}_w^{\text{global}}\mathbf{w}$. The proof problem is to finding a lower bound for $J_{\text{global}}\left(\mathbf{w}_c^{\text{corrected}}\right)$. The slack condition gives $B = \rho W - \epsilon$, and thus the objective function can be simplified:

$$J_{\text{global}}\left(\mathbf{w}_c^{\text{corrected}}\right) = \frac{B}{W} = \rho - \frac{\epsilon}{W}$$

The problem therefore reduces to finding a lower bound for $W = \mathbf{w}^T\mathbf{S}_w^{\text{global}}\mathbf{w}$. Assuming $\mathbf{S}_w^{\text{global}}$ is positive definite, there exists a $\mu_{\min}$ such that $\mathbf{S}_w^{\text{global}} \succeq \mu_{\min} I$. Consequently, $W \geqslant \mu_{\min}||\mathbf{w}_c^{\text{corrected}}||^2$. From the stationarity condition of the Lagrangian, we have:

$$\nabla_{\mathbf{w}}\mathcal{L} = \mathbf{w}_{\text{c}}^{\text{corrected}} - \mathbf{w}_{\text{c}}^{\text{local}}) -$$
$$2\lambda(\rho\mathbf{S}_w^{\text{global}}\mathbf{w}_{\text{c}}^{\text{corrected}} - \mathbf{S}_b^{\text{global}}\mathbf{w}_{\text{c}}^{\text{corrected}}) = 0$$

Transposing $\mathbf{w}_c^{local}$ and left-multiplying by $(\mathbf{w}_c^{corrected})^T$ yields:

$$(\mathbf{w}_c^{\text{corrected}})^T(\mathbf{w}_c^{\text{corrected}} - \mathbf{w}_c^{\text{local}}) =$$
$$2\lambda(\mathbf{w}_c^{\text{corrected}})^T(\mathbf{S}_b^{\text{global}} - \rho\mathbf{S}_w^{\text{global}})\mathbf{w}_c^{\text{corrected}}$$

Applying the complementary slackness condition $\lambda(\rho W - B - \epsilon) = 0$, the right-hand side can be simplified. Utilising the square-of-norm expansion for the vector difference, we obtain:

$$||\mathbf{w}_c^{\text{corrected}} - \mathbf{w}_c^{\text{local}}||^2 + ||\mathbf{w}_c^{\text{corrected}}||^2 = ||\mathbf{w}_c^{\text{local}}||^2 - 4\lambda^*\epsilon$$

Since the left-hand side is non-negative, it follows that: $\lambda^* \leq \frac{\|\mathbf{w}_c^{\text{local}}\|^2}{4\epsilon}$ and $\|\mathbf{w}_c^{corrected}\|^2 \leqslant \|\mathbf{w}_c^{local}\|^2 - 4\lambda^*\epsilon$. It is obvious that when λ^* attains its upper bound, $\|\mathbf{w}_c^{corrected}\|^2$ has its upper bound to be zero. Since λ^* typically does not attain its upper bound, $||\mathbf{w}_c^{corrected}||^2$ can be greater than zero, implying that W possesses a positive lower bound. Therefore, as ϵ tends to zero, $J_global(\mathbf{w}_c^{\text{corrected}}) \geqslant \rho$ clearly holds.

Theorem 4 (Stability of the Performance Lower Bound). Assume that when $\epsilon = 0$, the optimisation problem has a unique solution $\mathbf{w}_0$, and the corresponding $W_0 = \mathbf{w}_0^T\mathbf{S}_w^{\text{global}}\mathbf{w}_0 > 0$. Then, there exist constants $\epsilon_0 > 0$ and $c > 0$ such that for all $0 < \epsilon < \epsilon_0$,we have:

$$W = (\mathbf{w}_c^{\text{corrected}})^T\mathbf{S}_w^{\text{global}}\mathbf{w}_c^{\text{corrected}} \geq c$$

Local Non-Linear Search Strategy

The generalised Rayleigh quotient is founded upon two core linear hypotheses: namely, linear separability, and the assumption that data from different classes follow a Gaussian distribution with homoscedasticity (i.e., similar covariance matrices across classes). Consequently, when the data originates from a non-linear manifold $\mathcal{M}$ exhibiting significant local curvature variations, the local metric structure $g_{\lambda\iota}$ within different regions $U \subset \mathcal{M}$ of the manifold more closely approximates the conditions of the generalised Rayleigh quotient hypothesis. Furthermore, the strategy of computing weight vectors separately for each class (by partitioning data into "the class" and"not the class") smooths out inter-class differences, making the condition of equal covariance matrix of each class more likely to be satisfied. Therefore, we abandon the search for a global optimum and instead, via multi-objective evaluation using a genetic algorithm, seek a"high-quality local region $\mathcal{N}^*(\mathbf{x}_0)$ sufficiently representative of the test point. This approach can yield a superior discriminant boundary compared to global holistic discriminant trends. We compare two distinct search criteria: assessment by distance similarity and assessment by bidirected neighbour saliency.

Assessment by Distance Similarity

This genetic strategy aims to find a local region that is both spatially compact around the test point and exhibits an internal class distribution highly similar to the global data distribution. A neighbourhood centred on the test point $\mathbf{x}_0$ with radius r is defined as:

$$\mathcal{N}_r(\mathbf{x}_0) = \{\mathbf{x}_i \in \mathcal{D} \mid d(\mathbf{x}_0, \mathbf{x}_i) \leq r\}$$

where where $d(\cdot, \cdot)$ employs the Mahalanobis distance metric. The joint optimisation objectives are:

$$F(r) = \exp\left(-\gamma \cdot \tfrac{1}{|\mathcal{N}|}\textstyle\sum_{\mathbf{x},\in\mathcal{N}} \mathrm{d}(\mathbf{x}_0, \mathbf{x}_i)\right)$$
$$+\lambda \cdot \left[1 - D_{\text{KL}}\left(P_C^{\text{global}} \parallel P_C^{\mathcal{N}_r}\right)\right]$$

Assessment by bidirectional neighbour saliency

This strategy constructs a local region $\mathcal{N}^*(\mathbf{x}_0)$ for the test point $\mathbf{x}_0$ by synthesising geometric proximity and category representativeness through a multi-step smoothing technique. The core idea is to compute a distance adjustment factor for each candidate sample, effectively quantifying and integrating its class representativeness into the distance metric. Step 1: For each class C_i, select the m samples closest to $\mathbf{x}_0$ (based on original distance) to form a class-specific candidate set $\mathcal{S}_c$. Step 2: Compute the tightness of each class, defined as the reciprocal of the average distance of samples in $\mathcal{S}_c$ to their centroid $\boldsymbol{\mu}_c$, serving as a measure of class cohesion and obtain the relative tightness R_c:

$$tightness(c) = \frac{1}{\frac{1}{|\mathcal{S}_c|}\sum_{\mathbf{x}_i \in \mathcal{S}_c} |\mathbf{x}_i - \boldsymbol{\mu}_c| + \epsilon}$$

$$R_c = \frac{tightness(c)}{\frac{1}{C}\sum_{k=1}^{C} tightness(k)}$$

Step 3: Cascade Smoothing for Representativeness Assessment For each sample $\mathbf{x}_i$ in S_c, a representativeness score is computed to adjust its distance to $\mathbf{x}_0$. 1) Backward Ranking: Calculate the distance rank of $\mathbf{x}_0$ within S_c when using $\mathbf{x}_i$ as the reference point and normalise this rank to $[0, 1]$:

$$r_i = 1 - \frac{\text{backward}_{\text{rank}}(\mathbf{x}_i, \mathbf{x}_{\text{target}}, \mathcal{S}_c) - 1}{|\mathcal{S}_c| - 1}$$

2) Bayesian Smoothing: To stabilise the rank estimate under small sample conditions, the observed rank r_i is shrunk towards a prior using a Beta-binomial model (for example, with a uniform prior):

$$p_{\text{bayes}} = \frac{r_i \cdot N + \alpha}{N + \alpha + \beta}$$

3) Wilson Score Interval Smoothing: A conservative lower bound of the representativeness is computed via the Wilson score interval to account for estimation uncertainty, where z is the standard normal quantile (e.g., 1.96 for 95% confidence):

$$p_{\text{wilson}} = \frac{p_{\text{bayes}} + \frac{z^2}{2N} - z\sqrt{\frac{p_{\text{bayes}}(1-p_{\text{bayes}})}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}$$

4) The Wilson estimate is adjusted by the relative class tightness, enhancing the score for tight clusters ($R_c > 1$) and reducing it for loose ones ($R_c < 1$):

$$p_{\text{weighted}} = p_{\text{wilson}} \cdot (1 + \eta \cdot \tanh(R_c - 1))$$

Step 4: The final smooth adjustment factor is obtained by mapping a penalty term (e.g., $penalty = 1 - p_{weighted}$) through a Gaussian kernel. This ensures samples with high representativeness have their distances significantly reduced.

$$\lambda_i = \exp\left(-\frac{(penalty)^2}{2\sigma^2}\right)$$

Step 5: The final adjusted distance for sample $\mathbf{x}_i$ to the test point is:

$$d_{\text{adjusted}}(\mathbf{x}_i, \mathbf{x}_{target}) = d_{\text{original}}(\mathbf{x}_i, \mathbf{x}_{target}) \times \lambda_i$$

Proposed Algorithm

Step 1: To identify a high-quality neighbourhood $\mathcal{N}^*(\mathbf{x}_0)$ proximate to the test point $\mathbf{x}_0$ that is representative of the local manifold structure.

Step 2: Within the optimal local region $\mathcal{N}^*(\mathbf{x}_0)$, for each competing class $c \in C$, compute a local discriminant weight vector by maximising the generalised Rayleigh quotient (Fisher's criterion).

$$\mathbf{w}_{\text{local}}^{(c)} = \arg\max_{\mathbf{w} \in \mathbb{R}^d} \frac{\mathbf{w}^T \mathbf{S}_b^{(c)} \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w^{(c)} \mathbf{w}}$$

The closed-form solution is $\mathbf{w}_{\text{local}}^{(c)} = (\mathbf{S}_w^{(c)})^{-1}(\boldsymbol{\mu}_+^{(c)} - \boldsymbol{\mu}_-^{(c)})$, which is then L2-normalised.

Step 3: Manifold-Aware Refinement of Local Weights The local weight vectors are refined, by solving an optimisation problem that trades off fidelity to the local solution and the discriminant power on the global manifold.

$$\mathcal{F}(\mathbf{w}) = \alpha \cdot |\mathbf{w} - \mathbf{w}_{\text{local}}|^2 - (1 - \alpha) \cdot \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w} + \epsilon}$$

The final weight $\mathbf{w}_{\text{final}}^{(c)}$ is obtained via gradient-based optimization, $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta_t \cdot \nabla \mathcal{F}(\mathbf{w}^{(t)})$, with the search space constrained to $\mathbf{w} \in [-1, 1]^d$.

Step 4: Category-Aware Weighted k-Nearest Neighbour Search

For the final classification hypothesis for each $c \in C$, compute a category-specific weighted distance for all training samples using the corresponding refined weight vector $\mathbf{w}_{\text{final}}^{(c)}$, and select the k nearest neighbours $\mathcal{N}_k^{(c)}$.

$$d^{(c)}(\mathbf{x}_0, \mathbf{x}_i) = \sqrt{\sum_{j=1}^{d} w_{\text{final},j}^{(c)} \cdot (\mathbf{x}_{0,j} - \mathbf{x}_{i,j})^2}$$

Step 5: Decision Based on Homogeneity and Distance-Weighted Voting

For each class c, a final score is computed by combining the frequency of class c in its specific k-neighbour set $\mathcal{N}_k^{(c)}$ with a distance-weighted vote.

$$V_{\text{freq}}^{(c)} = \frac{|\{\mathbf{x}_i \in \mathcal{N}_k^{(c)} : y_i = c\}|}{k}$$

$$S(c) = V_{\text{freq}}^{(c)} \cdot V_{\text{dist}}^{(c)}$$

The final predicted class is:

$$\hat{y} = \arg\max_{c \in C} S(c)$$

Experiment

This section aims to conduct a rigorous comparative experiment to evaluate the overall performance of the proposed Manifold-Corrected Adaptive Weighted K-Nearest Neighbours (MCAW-KNN) algorithm. The experiment is designed to address two core questions: (1) Whether the classification accuracy of the proposed algorithm is competitive compared to mainstream classical classification algorithms; (2) Whether the weight learning strategy of "local adaptive weighting and global manifold correction" adopted in KNN is superior compared to other feature weighting schemes.

The experiment encompasses two main comparative categories. Firstly, the proposed algorithm will be compared against six representative mainstream algorithms as benchmarks: Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, Gradient Boosting Tree and Naive Bayes. Secondly, a comparison will be made against various KNN weighting schemes to evaluate the proposed weighting strategy. These methods include KNN with Random Forest weights (utilising feature importance from Random Forest), KNN with PCA weights, KNN with Pearson correlation coefficient, KNN with Spearman's rank correlation coefficient, KNN with Gini index, KNN with ReliefF analysis weights

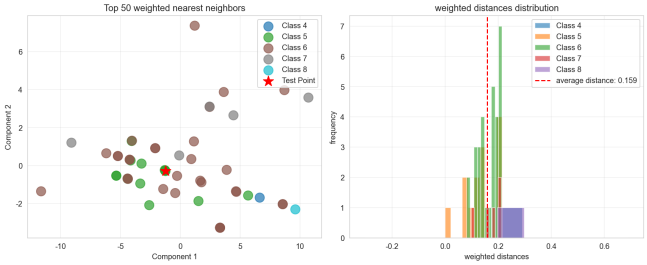

Figure 1: the nearest sample overlapped

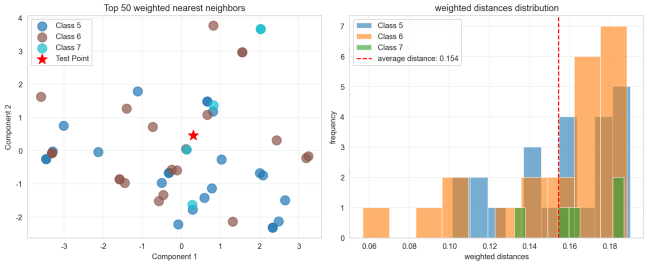

Figure 2: the ambiguously proximate samples

(evaluating feature distinguishing power via the ReliefF algorithm), KNN with Lasso regression, KNN with mutual information weights (based on mutual information between features and the target variable), and KNN with entropy weighting.

To ensure a comprehensive evaluation, all experiments will employ a five-repeated hold-out method with different random seeds. The core evaluation metric is the classification accuracy, for which the standard deviation will also be calculated

to assess stability. The experiments utilise two types of datasets with distinct characteristics: firstly, classical linearly separable datasets, which feature balanced class distributions and a high degree of linear separability, serving to validate the fundamental effectiveness of the algorithms; and secondly, datasets characterised by imbalance and high overlap in attribute value distributions. The inherent class overlap in this second category of datasets tends to obscure classification boundaries, thereby posing greater demands on the robustness of the algorithms. In this experiment, these datasets will be used without any preprocessing specifically targeting noise, outliers, or class imbalance, in order to directly test the raw performance of each algorithm in complex, noisy real-world scenarios.

As illustrated in Figure 1, following a two-dimensional Principal Component Analysis (PCA) transformation, the nearest point to the test instance (where a class 5 sample overlaps with the test sample) remains the closest in terms of the weighted distance after applying a local neighbourhood weighting scheme. In contrast, Figure 2 demonstrates a scenario where the nearest samples to the test point is ambiguous under the PCA transformation. However, by extracting a high-quality neighbourhood around the test point, the method successfully attracts proximate samples of the genuine class (class 5) whilst repelling nearby samples from other classes. Consequently, a distinct and discernible decision boundary is effectively generated in the vicinity of the test point.

The results of table 1 demonstrate that the performance of all algorithms is significantly influenced by dataset characteristics. The proposed MCAW-KNN algorithm achieves the highest accuracy on the Iris, network intrusion detection detection datasets, whilst MCAW-KNN(GA) performs best on the more complex winequalityN and network intrusion datasets. On the classic wine dataset, traditional algorithms like Random Forest, SVM, and Logistic Regression yield the top results.

The differential performance between MCAW-KNN(GA) and MCAW-KNN offers insights into their respective mechanisms. MCAW-KNN(GA), leveraging a genetic algorithm, excels in complex scenarios (i.e., winequalityN) where defining a universally good neighbourhood is difficult. The GA effectively performs a global optimisation to find a neighbourhood structure that best supports the subsequent manifold correction. In contrast, the standard MCAW-KNN performs exceptionally well on datasets where local structure is more reliable (i.e., Iris, network intrusion detection detection), as its direct use of reverse nearest neighbours and class compactness efficiently amplifies local class signals without the overhead of a global search.

A limitation observed is the performance on the wine dataset, where simpler models prevailed. This indicates that the proposed method's overhead may not be justified for straightforward classification problems. Future work could investigate a preliminary data complexity analysis to decide whether to employ the full manifold correction procedure or default to a simpler classifier.

As shown in Table 2, the two algorithms proposed in this paper demonstrate excellent or highly competitive classification performance on most datasets. On the Iris dataset, the MCAW-KNN algorithm based on reverse nearest neighbours achieved the best performance amongst all methods, with an average accuracy of 0.9777 and the lowest standard deviation (±0.0157), showcasing exceptional accuracy and stability. The performance of MCAW-KNN(GA) (0.9533) was on par with that of most traditional methods. This result provides strong evidence that, on datasets with compact class structures, constructing a local region can more precisely capture the essential characteristics of a class, thereby leading to reliable classification decisions.

Notably, the advantage of the new algorithms is most pronounced on the "winequalityN" and "network intrusion detection" datasets, which are characterised by higher complexity and potential class distribution imbalance. In contrast, conventional methods such as KNN+RF, KNN+Gini, and KNN+Lasso attained the highest average accuracy (0.9944) on the Wine dataset. The accuracy of the proposed MCAW-

	iris	wine	winequalityN	network intrusion
Random Forest	0.9333 (±0.0422)	0.9833 (±0.0136)	0.7576 (±0.0429)	0.9922 (±0.0157)
Decision Tree	0.9067 (±0.0573)	0.8944 (±0.0369)	0.6545 (±0.0653)	0.9922 (±0.0157)
SVM	0.9600 (±0.0327)	0.9833 (±0.0136)	0.5758 (±0.0939)	0.9882 (±0.0157)
Logistic Regression	0.9400 (±0.0389)	0.9833 (±0.0136)	0.5394 (±0.0445)	0.9490 (±0.0342)
Gradient Boosting Tree	0.9000 (±0.0667)	0.9444 (±0.0351)	0.5939 (±0.0680)	0.9922 (±0.0157)
Naive Bayes	0.9333 (±0.0422)	0.9778 (±0.0111)	0.3030 (±0.0742)	0.8784 (±0.0560)
AdaBoost	0.9400 (±0.0389)	0.9556 (±0.0377)	0.4667 (±0.0624)	0.9843 (±0.0229)
MCAW-KNN(GA)	0.9533 (±0.0339)	0.88329 (±0.0408)	0.8060 (±0.0804)	0.9922 (±0.0096)
MCAW-KNN	0.9777 (±0.0157)	0.8819 (±0.0455)	0.7697 (±0.0562)	0.9961 (±0.0078)

Table 1: Comparison with mainstream classifiers

	wine	iris	winequalityN	network intrusion
MCAW-KNN	0.8819(±0.0455)	0.9777 (±0.0157)	0.7697 (±0.0562)	0.996078 (±0.0078)
MCAW-KNN(GA)	0.88329(±0.0408)	0.9533 (±0.0339)	0.8060 (±0.0804)	0.99215 (±0.0096)
KNN+RF	0.9889 (±0.0136)	0.9533 (±0.0340)	0.7212 (±0.0618)	0.9889 (±0.0063)
KNN+PCA	0.9833 (±0.0136)	0.9400 (±0.0389)	0.6788 (±0.0492)	0.9905 (±0.0078)
KNN+ Pearson	0.9833 (±0.0136)	0.9733 (±0.0133)	0.7091 (±0.0781)	0.9873 (±0.0108)
KNN+ Spearman's	0.9778 (±0.0208)	0.9667 (±0.0000)	0.7030 (±0.0588)	0.9841 (±0.0123)
KNN+Gini	0.9889 (±0.0136)	0.9533 (±0.0340)	0.7212 (±0.0618)	0.9889 (±0.0063)
KNN+Lasso	0.9833 (±0.0136)	0.9533 (±0.0340)	0.7152 (±0.0364)	0.9905 (±0.0078)
KNN+ReliefF	0.9778 (±0.0111)	0.9533 (±0.0340)	0.7091 (±0.0309)	0.9905 (±0.0117)
KNN+MI	0.9889 (±0.0136)	0.9733 (±0.0133)	0.6909 (±0.0702)	0.9841 (±0.0123)
KNN+entropy	0.9722 (±0.0176)	0.9533 (±0.0340)	0.7030 (±0.0588)	0.9905 (±0.0117)

Table 2: Comparison with mainstream weighting schemes

KNN(GA) and MCAW-KNN is slightly lower, at 0.8833 and 0.8819 respectively; however, their standard deviations (±0.0408, ±0.0455) are smaller than those of most comparative methods. This indicates that whilst the average performance of our method is marginally inferior, its prediction outcomes are more stable. This suggests that on datasets with a clear feature-space structure and good inter-class separation, the marginal benefit of fine-grained local adaptive weighting may be less than the computational complexity it introduces, yet the robustness of the algorithm is preserved.

Conclusion

The performance of the two proposed algorithms is closely tied to dataset complexity. On the Iris and network intrusion detection datasets with compact class structures, the MCAW-KNN algorithm based on statistical representativeness evaluation achieved optimal accuracy and stability. Conversely, on the highly complex and imbalanced winequalityN dataset, the MCAW-KNN (GA) method, which employs a global search via genetic algorithms, significantly outperformed all traditional methods. For the Wine dataset, which exhibits good linear separability, traditional global weighting methods proved sufficiently effective. The proposed algorithms, while incurring a notable accuracy cost due to their more localised nature, offered more stable predictions. This validates the initial aim of the proposed algorithms—to handle complex data distributions through local adaptive weighting—with their advantages becoming pronounced when the problem complexity is sufficiently high.

References

[1] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[2] Y. Gao and Y. Liu. (2016). An improved feature-weighted method based on K-NN. 2016 35th Chinese Control Conference (CCC), pp. 6950-6956

[3] Çetin, A. İ., & Büyüklü, A. H. (2025). A new approach to K-nearest neighbors distance metrics on sovereign country credit rating. Kuwait Journal of Science, 52(1), 100324.

[4] Zhang, C., Yao, J., Hu, G., & Schott, T. (n.d.). Applying Feature-Weighted Gradient Decent K-Nearest Neighbor to Select Promising Projects for Scientific Funding.

[5] Biswas, Nimagna & Chakraborty, Saurajit & Mullick, Sankha & Das, Swagatam. (2017). A Parameter Independent Fuzzy Weighted k -Nearest Neighbor Classifier. Pattern Recognition Letters. 101. 10.1016/j.patrec.2017.11.003.

[6] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.

[7] Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics, 14(1), 106.

[8] López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced

data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences, 250, 113–14
[9] García, V., Sánchez, J. S., & Mollineda, R. A. (2012). On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowledge-Based Systems, 25, 13–21.
[10] Napierala, K., & Stefanowski, J. (2016). Types of minority class examples and their influence on learning classifiers from imbalanced data. Journal of Intelligent Information Systems, 46(3), 563–597.
[11] Roweis, S.T., & Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290 5500, 2323-6 .
[12] Spiers, R., & Kalivas, J. H. (2025) Local Adaptive Fusion Regression (LAFR) for Local Linear Multivariate Calibration: Application to Large Datasets. Appl Spectrosc. 2025 May;79(5):797-807.
[13] Ahmad, T., Portier, F., & Stupfler, G. (2024) Local logistic regression for dimension reduction in classification arXiv
[14] Holland, J. H. (1975). Adaptation in Natural and Artificial Systems MIT Press, Cambridge, MA.
[15] De Jong, K. A. (1975) An Analysis of the Behavior of a Class of Genetic Adaptive Systems Ph.D. Thesis, University of Michigan, Ann Arbor, MI.
[16] Prasetio, R.T. (2020). Genetic Algorithm to Optimize k-Nearest Neighbor Parameter for Benchmarked Medical Datasets Classification.
[17] Shen, Y., Wu, C., Luo, J., & Gao, F. (2023). KNN Intrusion Detection Model Based on Meta-optimization. Journal of Beijing University of Technology.
[18] Han, B., Qiao, L.-N., Chen, J.-L., Zhang, X.-D., Zhang, Y.-X., & Zhao, Y.-H. (2021). GeneticKNN: a weighted KNN approach supported by genetic algorithm for photometric redshift estimation of quasars. Publications of the Astronomical Society of the Pacific, 133(1020), 1-12.
[19] García-Gutiérrez, J., Mateos-García, D., & Muñoz-Pérez, J. (2019). On the evolutionary weighting of neighbours and features in the k-nearest neighbour rule. Neurocomputing, 353, 25-40.
[20] Mateos-García, Daniel & García-Gutiérrez, Jorge & Riquelme, José. (2012). On the evolutionary optimization of k-NN by label-dependent feature weighting. Pattern Recognition Letters. 33. 10.1016/j.patrec.2012.08.011.
[21] McRoberts, R.E., Domke, G.M., Chen, Q., Næsset, E., & Gobakken, T. (2016). Using genetic algorithms to optimize k-Nearest Neighbors configurations for use with airborne laser scanning data.
[22] Kahraman, H. T., Bayindir, R., & Sagiroglu, S. (2012). A new approach to predict the excitation current and parameter weightings of synchronous machines based on genetic algorithm-based k-NN estimator. Energy Conversion and Management, 64, 129–138.
[23] Waller, L. A., & Gotway, C. A. (2004). Applied Spatial Statistics for Public Health Data. John Wiley & Sons.
[24] Chen, C., Wakefield, J., & Lumley, T. (2014). The use of sampling weights in Bayesian hierarchical models for small area estimation. Spatial and Spatio-temporal Epidemiology, 11, 33-43.
[25] Lemke, D., Mattauch, V., Heidinger, O. et al. (2013). Detecting cancer clusters in a regional population with local cluster tests and Bayesian smoothing methods: a simulation study. Int J Health Geogr 12, 54.
[26] Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician, 52(2), 119-126.
[27] Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. Statistical Science, 16(2), 101-133.
[28] Belkin, M., Niyogi, P., & Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. Journal of Machine Learning Research, 7(Nov), 2399-2434.
[29] Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research, 10(Feb), 207-244.
[30] Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition (2nd ed.). Academic Press.
[31] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", 986-996.