Université Mohammed V - Rabat Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes

Project 2:

Comparative study on different types of kernels

Group:

Diani Achraf
Daoudi Mohammed Oualid

Supervised by:

Mr El Afia Abdellatif
Mr Hssayni El Houssaine

Année Universitaire 2023-2024

Contents

List of Figures

Chapter 1

Introduction

In machine learning, encountering non-linearly separable data poses a significant challenge for classification tasks. When traditional linear classifiers like Support Vector Machines (SVMs) struggle with such data, the concept of kernels comes into play. Kernels offer a solution by transforming the original feature space into a higher-dimensional space where the data becomes separable. By adding new features derived from the existing ones, kernels enable the creation of decision boundaries that can effectively classify non-linearly separable data. This approach extends the applicability of classifiers like SVMs to a wider range of datasets, making them more versatile and powerful in practical applications.

1.1 Kernel Trick

1.1.1 Definition

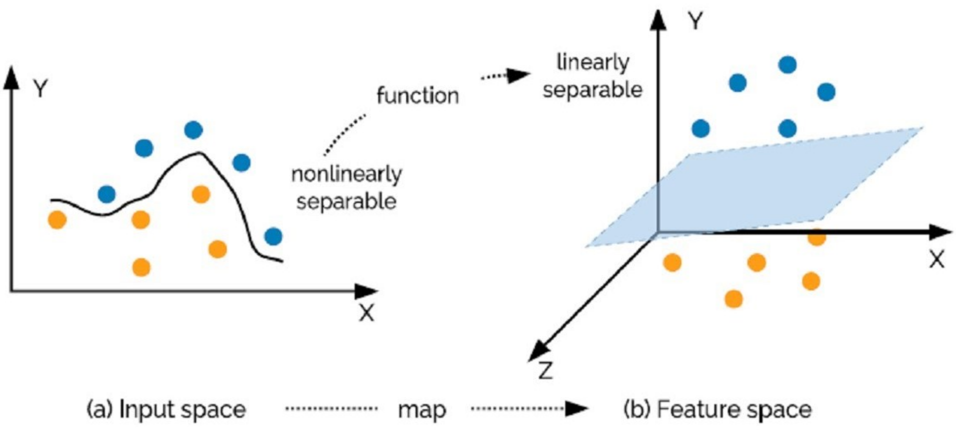

Figure 1.1: Fitness comparison between the three algorithms

Kernel methods in machine learning are algorithms designed to analyze and model complex data effectively. They achieve this by mapping input data into a higher-dimensional feature space, facilitating easier separation or analysis. Simply put, kernel methods transform data into a more conducive space for identifying patterns and relationships. This transformation is enabled by a kernel function, which computes similarity measures between pairs of data points, enhancing the algorithm's ability to capture intricate structures within the data.

A function that takes vectors in the original space as its inputs and returns the dot product of the vectors in the feature space is called a kernel function. More formally, if we have data $x, z \in X$ and a map $\psi : X \to \mathbb{R}^n$, then $K(x, z) = \langle \psi(x), \psi(z) \rangle$ is a kernel function. Our kernel function accepts inputs in the original lower-dimensional space and returns the dot product of the transformed vectors in the higher-dimensional space.

1.1.2 Feature Maps

We start by considering fitting cubic functions $y = \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$. It turns out that we can view the cubic function as a linear function over a different set of feature variables. Concretely, let the function $\phi : \mathbb{R} \to \mathbb{R}^4$ be defined as

$$\phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} \in \mathbb{R}^4$$

where x represents the input attribute, and $\phi(x)$ represents the feature variables. Let $\theta \in \mathbb{R}^4$ be the vector containing $\theta_0, \theta_1, \theta_2, \theta_3$ as entries. Then we can rewrite the cubic function in x as:

$$\theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0 = \theta^T \phi(x)$$

Thus, a cubic function of the variable x can be viewed as a linear function over the variables $\phi(x)$.

1.2 Types of Kernels

1.2.1 Linear Kernel

The linear kernel is the simplest type of kernel function. It is defined as the dot product of two vectors:

$$K(x, y) = x^T y$$

where x and y are vectors. The linear kernel is effective when the data is linearly separable, meaning that a single straight line (or hyperplane in higher dimensions) can separate the classes.

1.2.2 Non-linear Kernels

Non-linear kernels allow for more complex decision boundaries that can capture non-linear relationships in the data. These kernels are particularly useful when the data is not linearly separable. Common non-linear kernels include polynomial and Gaussian (RBF) kernels.

Polynomial Kernel

The polynomial kernel is defined as:

$$K(x, y) = (\alpha x^T y + c)^d$$

where α, c, and d are parameters. This kernel allows for polynomial decision boundaries, making it useful for capturing polynomial relationships between features.

Gaussian (RBF) Kernel

The Gaussian kernel, also known as the Radial Basis Function (RBF) kernel, is defined as:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

where σ is a parameter that determines the width of the Gaussian function. The RBF kernel can capture very complex relationships by mapping the data into an infinite-dimensional space.

Sigmoid Kernel

The sigmoid kernel is defined as:

$$K(x, y) = \tanh(\alpha x^T y + c)$$

where α and c are parameters. This kernel is inspired by the activation function of neural networks and can model relationships that resemble those captured by neural networks.

1.3 Building Kernels with Mercer's Theorem

Mercer's Theorem provides a foundational criterion for a function to be a valid kernel. Specifically, a function $\kappa(x, x')$ is a Mercer kernel if it results in a positive semidefinite kernel matrix for any finite set of input data points.

1.3.1 Steps to Build a Kernel

Define Symmetric Function

Start with a symmetric function $\kappa(x, x')$:

$$\kappa(x, x') = \kappa(x', x)$$

Ensure Positive Semidefiniteness

The function must ensure that for any finite set of points $\{x_1, x_2, \ldots, x_n\}$, the kernel matrix K, where $K_{ij} = \kappa(x_i, x_j)$, is positive semidefinite. This means:

$$c^T K c \geq 0$$

for any vector $c \in \mathbb{R}^n$.

Combine Kernels

Combine basic kernels to create new ones, ensuring the resultant function remains a valid Mercer kernel:

- **Addition**: If κ_1 and κ_2 are Mercer kernels, then $\kappa(x, x') = \kappa_1(x, x') + \kappa_2(x, x')$ is also a Mercer kernel.
- **Multiplication**: If κ_1 and κ_2 are Mercer kernels, then $\kappa(x, x') = \kappa_1(x, x') \cdot \kappa_2(x, x')$ is also a Mercer kernel.
- **Scaling**: If κ is a Mercer kernel and $\alpha > 0$, then $\alpha\kappa(x, x')$ is also a Mercer kernel.

Verify Positive Semidefiniteness Practically

For practical purposes, compute the kernel matrix K for a sample dataset and verify that all eigenvalues of K are non-negative.

Chapter 2

Detailed Kernel Methods

2.1 Simple Kernel

In the expansive landscape of machine learning and kernel methods, the Simple Kernel, typically denoted as $K(x, y)$, serves as a fundamental mathematical tool with a diverse range of applications. This kernel is formulated as the inner product of two input vectors x and y. Symbolically, this can be expressed as:

$$K(x, y) = \langle x, y \rangle$$

The simplicity of the Simple Kernel makes it a versatile choice in several machine learning scenarios, spanning from classification to regression, and more.

2.1.1 Kernel Demonstration

To prove that the Simple Kernel is a valid kernel, we need to show that the corresponding kernel function, $K(x, y)$, satisfies the properties of symmetry and positive definiteness.

Symmetry

To prove symmetry, we need to show that $K(x, y) = K(y, x)$ for any input vectors x and y.

$$K(x, y) = \langle x, y \rangle, \quad K(y, x) = \langle y, x \rangle$$

We can observe that $\langle x, y \rangle = \langle y, x \rangle$ since the inner product is symmetric. Thus, we have:

$$K(x, y) = \langle x, y \rangle = \langle y, x \rangle = K(y, x)$$

This proves symmetry.

Positive Definiteness

For any finite set of data points $\{x_1, x_2, \ldots, x_n\}$, the kernel matrix K can be defined as:

$$K_{ij} = K(x_i, x_j) = \langle x_i, x_j \rangle$$

To prove positive definiteness, we need to show that for any vector $c = [c_1, c_2, \ldots, c_n]$, the quadratic form $c^T K c \geq 0$.

$$c^T K c = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \langle x_i, x_j \rangle = \left\langle \sum_{i=1}^{n} c_i x_i, \sum_{j=1}^{n} c_j x_j \right\rangle = \left\| \sum_{i=1}^{n} c_i x_i \right\|^2 \geq 0$$

Thus, the Simple Kernel is positive definite.

2.1.2 Properties

The properties of the Simple Kernel are as follows:

1. The kernel is expressed as $K(x, y) = \langle x, y \rangle$.
2. The kernel value is unbounded.
3. The kernel function is continuous for any input vectors x and y.
4. The kernel value increases as the similarity between the input vectors x and y increases.

2.1.3 Problems Solved

The Simple Kernel excels in modeling linear data relationships. We'll now highlight problems effectively tackled using the Simple Kernel compared to other kernels:

1. **Linearity Problem:** The kernel is ideal for situations where relationships between variables are linear.
2. **Simple Class Separability:** The kernel works well when classes are linearly separable.

2.1.4 Applications of the Kernel

The Simple Kernel, with its capability to model linear relationships, has found use in various domains and scenarios in machine learning:

- **Support Vector Machines (SVM):** For linear classification.
- **Linear Regression:** For predicting continuous outputs.
- **Principal Component Analysis (PCA):** For dimensionality reduction.

2.2 Polynomial Kernel

The Polynomial Kernel extends the capabilities of the Simple Kernel by mapping input vectors into a higher-dimensional space, allowing it to capture non-linear relationships.

2.2.1 Kernel Definition

The Polynomial Kernel is defined as:

$$K(x, y) = (\langle x, y \rangle + c)^d$$

where c is a free parameter trading off the influence of higher-order versus lower-order terms, and d is the degree of the polynomial.

2.2.2 Kernel Demonstration

To prove that the Polynomial Kernel is a valid kernel, we need to show that the corresponding kernel function, $K(x, y)$, satisfies the properties of symmetry and positive definiteness.

Symmetry

To prove symmetry, we need to show that $K(x, y) = K(y, x)$ for any input vectors x and y.

$$K(x, y) = (\langle x, y\rangle + c)^d, \quad K(y, x) = (\langle y, x\rangle + c)^d$$

We can observe that $\langle x, y\rangle = \langle y, x\rangle$ since the inner product is symmetric. Thus, we have:

$$K(x, y) = (\langle x, y\rangle + c)^d = (\langle y, x\rangle + c)^d = K(y, x)$$

This proves symmetry.

Positive Definiteness

For any finite set of data points $\{x_1, x_2, \dots, x_n\}$, the kernel matrix K can be defined as:

$$K_{ij} = K(x_i, x_j) = (\langle x_i, x_j\rangle + c)^d$$

To prove positive definiteness, we need to show that for any vector $c = [c_1, c_2, \dots, c_n]$, the quadratic form $c^T K c \geq 0$.

$$c^T K c = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j (\langle x_i, x_j\rangle + c)^d$$

Since polynomial kernels represent valid inner products in a higher-dimensional space, they are positive definite.

2.2.3 Properties

The properties of the Polynomial Kernel are as follows:

1. The kernel is expressed as $K(x, y) = (\langle x, y\rangle + c)^d$.
2. The kernel can capture polynomial relationships between variables.
3. The kernel value is unbounded.
4. The kernel function is continuous for any input vectors x and y.

2.2.4 Problems Solved

The Polynomial Kernel excels in modeling non-linear data relationships. We'll now highlight problems effectively tackled using the Polynomial Kernel compared to other kernels:

1. **Non-Linearity Problem:** The kernel helps in situations where relationships between variables are not straight lines.
2. **Poor Class Separability:** The kernel can help by mapping the data into a higher-dimensional space to improve separability.
3. **Interaction Effects:** The kernel can help capture these non-linear interactions.

2.2.5 Applications of the Kernel

The Polynomial Kernel, with its capability to model non-linear relationships, has found use in various domains and scenarios in machine learning:

- **Support Vector Machines (SVM):** For non-linear classification.

- **Kernel Ridge Regression:** For capturing non-linear relationships.
- **Clustering:** In algorithms like kernel k-means.

2.3 Gaussian Kernel

The Gaussian Kernel, also known as the Radial Basis Function (RBF) Kernel, is widely used in machine learning for its ability to handle the complexity of the data by mapping it into an infinite-dimensional space.

2.3.1 Kernel Definition

The Gaussian Kernel is defined as:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

where σ is a free parameter that controls the width of the Gaussian function.

2.3.2 Kernel Demonstration

To prove that the Gaussian Kernel is a valid kernel, we need to show that the corresponding kernel function, $K(x, y)$, satisfies the properties of symmetry and positive definiteness.

Symmetry

To prove symmetry, we need to show that $K(x, y) = K(y, x)$ for any input vectors x and y.

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right), \quad K(y, x) = \exp\left(-\frac{\|y - x\|^2}{2\sigma^2}\right)$$

We can observe that $\|x - y\|^2 = \|y - x\|^2$. Thus, we have:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\|y - x\|^2}{2\sigma^2}\right) = K(y, x)$$

This proves symmetry.

Positive Definiteness

For any finite set of data points $\{x_1, x_2, \ldots, x_n\}$, the kernel matrix K can be defined as:

$$K_{ij} = K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

To prove positive definiteness, we need to show that for any vector $c = [c_1, c_2, \ldots, c_n]$, the quadratic form $c^T K c \geq 0$.

$$c^T K c = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j K_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

Since the Gaussian Kernel is derived from the Gaussian function, which is positive definite, it is also positive definite.

2.3.3 Properties

The properties of the Gaussian Kernel are as follows:

1. The kernel is expressed as $K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$.
2. The kernel can capture complex, non-linear relationships.
3. The kernel value is bounded between 0 and 1.
4. The kernel function is continuous for any input vectors x and y.

2.3.4 Problems Solved

The Gaussian Kernel excels in detecting complex data relationships. We'll now highlight problems effectively tackled using the Gaussian Kernel compared to other kernels:

1. **Non-Linearity Problem:** The kernel helps in situations where relationships between variables are highly non-linear.
2. **Poor Class Separability:** The kernel can help by mapping the data into an infinite-dimensional space to improve separability.
3. **Outlier Robustness:** The kernel can mitigate the impact of outliers on the model's performance.

2.3.5 Applications of the Kernel

The Gaussian Kernel, with its capability to model complex, non-linear relationships, has found use in various domains and scenarios in machine learning:

- **Support Vector Machines (SVM):** For non-linear classification.
- **Kernel Density Estimation:** For estimating probability densities.
- **Clustering:** In algorithms like spectral clustering.

2.4 Sigmoid Kernel

The Sigmoid Kernel, also known as the Hyperbolic Tangent Kernel, is inspired by the activation function of neural networks. It is widely used in machine learning for its ability to model relationships in data.

2.4.1 Kernel Definition

The Sigmoid Kernel is defined as:

$$K(x, y) = \tanh(\alpha\langle x, y\rangle + c)$$

where α and c are kernel parameters.

2.4.2 Kernel Demonstration

To prove that the Sigmoid Kernel is a valid kernel, we need to show that the corresponding kernel function, $K(x, y)$, satisfies the properties of symmetry and positive definiteness.

Symmetry

To prove symmetry, we need to show that $K(x, y) = K(y, x)$ for any input vectors x and y.

$$K(x, y) = \tanh(\alpha\langle x, y\rangle + c), \quad K(y, x) = \tanh(\alpha\langle y, x\rangle + c)$$

We can observe that $\langle x, y\rangle = \langle y, x\rangle$ since the inner product is symmetric. Thus, we have:

$$K(x, y) = \tanh(\alpha\langle x, y\rangle + c) = \tanh(\alpha\langle y, x\rangle + c) = K(y, x)$$

This proves symmetry.

Positive Definiteness

The positive definiteness of the Sigmoid Kernel is not guaranteed for all parameter values. It behaves similarly to the Gaussian and Polynomial kernels but may not be positive definite for certain α and c values.

2.4.3 Properties

The properties of the Sigmoid Kernel are as follows:

1. The kernel is expressed as $K(x, y) = \tanh(\alpha\langle x, y\rangle + c)$.
2. The kernel can capture non-linear relationships.
3. The kernel value is bounded between -1 and 1.
4. The kernel function is continuous for any input vectors x and y.

2.4.4 Problems Solved

The Sigmoid Kernel excels in modeling relationships inspired by neural networks. We'll now highlight problems effectively tackled using the Sigmoid Kernel compared to other kernels:

1. **Non-Linearity Problem:** The kernel helps in situations where relationships between variables are non-linear.
2. **Neural Network Representation:** The kernel can simulate neural network behavior.

2.4.5 Applications of the Kernel

The Sigmoid Kernel, with its capability to model non-linear relationships, has found use in various domains and scenarios in machine learning:

- **Support Vector Machines (SVM):** For non-linear classification.
- **Neural Network Emulation:** For simulating neural network activation functions.
- **Clustering:** In algorithms like kernel k-means.

Chapter 3

Breast Cancer Dataset

In this chapter, we will discuss the breast cancer dataset, its significance in medical research, and the application of Support Vector Machines (SVM) using various kernels for classification tasks.

3.1 Dataset Overview

The breast cancer dataset is widely used in medical research to develop predictive models for breast cancer diagnosis. The dataset contains features derived from digitized images of fine needle aspirates (FNA) of breast masses. These features describe the characteristics of the cell nuclei present in the images.

3.1.1 Features

The dataset includes the following features:

- **ID Number:** Unique identifier for each patient.
- **Diagnosis:** Classification of the tumor (Malignant or Benign).
- **Radius:** Mean of distances from center to points on the perimeter.
- **Texture:** Standard deviation of gray-scale values.
- **Perimeter:** Perimeter of the tumor.
- **Area:** Area of the tumor.
- **Smoothness:** Local variation in radius lengths.
- **Compactness:** $\frac{\text{Perimeter}^2}{\text{Area}} - 1.0$
- **Concavity:** Severity of concave portions of the contour.
- **Concave Points:** Number of concave portions of the contour.
- **Symmetry:** Symmetry of the tumor.
- **Fractal Dimension:** $\frac{1}{\text{Area}}$ approximation.

3.2 Importance of the Dataset

Early and accurate diagnosis of breast cancer is crucial for effective treatment and improved survival rates. The breast cancer dataset provides valuable information that can be used to build machine learning models to assist in the diagnosis process. By analyzing the features of cell nuclei, these models can help distinguish between malignant and benign tumors, thus aiding in early detection and treatment.

3.3 Application of SVM with Kernels

Support Vector Machines (SVM) are powerful classifiers that can be used to create predictive models for breast cancer diagnosis. The choice of kernel plays a significant role in the performance of SVM. Here, we discuss the application of different kernels for the breast cancer dataset.

3.3.1 Simple Kernel

The Simple Kernel, or Linear Kernel, is effective when the data is linearly separable. In the context of breast cancer diagnosis, if the features of malignant and benign tumors are linearly separable, the Simple Kernel can be used to create a straightforward and efficient model.

3.3.2 Polynomial Kernel

The Polynomial Kernel can capture non-linear relationships between the features of the dataset. Breast cancer data may exhibit complex interactions between features, and the Polynomial Kernel can map these features into a higher-dimensional space, improving the classifier's ability to distinguish between malignant and benign tumors.

3.3.3 Gaussian Kernel

The Gaussian Kernel, or Radial Basis Function (RBF) Kernel, is well-suited for datasets with intricate patterns. It maps the data into an infinite-dimensional space, enabling the SVM to capture subtle variations in the features. This is particularly useful for breast cancer diagnosis, where the relationships between features may be highly non-linear.

3.3.4 Sigmoid Kernel

The Sigmoid Kernel, inspired by neural network activation functions, can model relationships that resemble those captured by neural networks. It provides another approach to handling non-linearities in the breast cancer dataset, offering flexibility in creating the classification model.

3.4 Advantages of Using SVM for Breast Cancer Diagnosis

The advantages of using SVM for breast cancer diagnosis include:

- **High Accuracy:** SVMs are known for their high classification accuracy, making them suitable for medical diagnosis.
- **Robustness to Overfitting:** SVMs use regularization parameters to avoid overfitting, which is critical when dealing with medical data.
- **Flexibility with Kernels:** The ability to choose different kernels allows SVMs to adapt to various types of data distributions and complexities.
- **Effective in High-Dimensional Spaces:** SVMs perform well in high-dimensional feature spaces, which is beneficial when working with datasets like the breast cancer dataset that have multiple features.

By leveraging the strengths of SVM and the appropriate kernel functions, we can build robust models for the early and accurate diagnosis of breast cancer, ultimately contributing to better patient outcomes and more effective treatments.

Chapter 4

Execution Results

In this chapter, we present the execution results of applying various kernel methods using Support Vector Machines (SVM) on the breast cancer dataset. The figure below shows the accuracy comparison across different kernel methods. The execution was in GoogleColab using only the cpu.

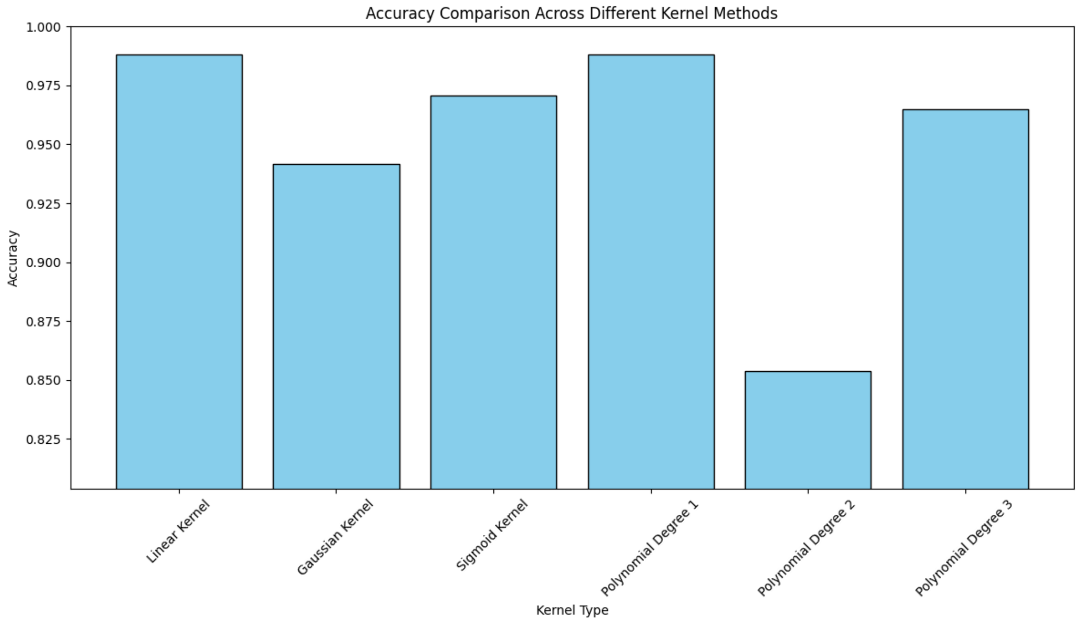

Figure 4.1: Fitness comparison between the three algorithms

4.1 Discussion

The accuracy comparison depicted in Figure demonstrates the performance of different kernel methods. Here, we discuss why certain kernels perform better than others:

4.1.1 Linear Kernel

The Linear Kernel achieves the highest accuracy among the tested kernels. This indicates that the breast cancer dataset is likely linearly separable, or close to it, making the linear decision boundary effective for classification. The simplicity and efficiency of the Linear Kernel contribute to its high performance in this scenario.

4.1.2 Polynomial Kernel

The Polynomial Kernel with degree 1, which is equivalent to the Linear Kernel, also shows high accuracy, consistent with the linear separability of the dataset. However, Polynomial kernels with higher degrees (2 and 3) show a significant drop in accuracy. This decrease in performance may be due to overfitting, where the higher-degree polynomials capture noise in the data rather than the underlying patterns. Overfitting leads to poor generalization on unseen data, thus reducing the model's accuracy.

4.1.3 Gaussian Kernel

The Gaussian Kernel, also known as the Radial Basis Function (RBF) Kernel, performs well, though slightly lower than the Linear and Polynomial degree 1 kernels. The Gaussian Kernel's ability to map data into an infinite-dimensional space allows it to capture non-linear relationships effectively. However, it requires careful tuning of the σ parameter to achieve optimal performance. The slight drop in accuracy may be due to suboptimal parameter tuning or the inherent characteristics of the dataset.

4.1.4 Sigmoid Kernel

The Sigmoid Kernel, inspired by neural network activation functions, shows good performance but is not as high as the Linear or Gaussian kernels. The Sigmoid Kernel can model non-linear relationships but may not be as effective for this dataset due to the specific characteristics of the data. Additionally, the Sigmoid Kernel's performance can be sensitive to the choice of its parameters, α and c, which may not have been optimally selected in this experiment.

4.2 Conclusion

The execution results highlight that the Linear Kernel and Polynomial Kernel with degree 1 achieve the highest accuracy on the breast cancer dataset. This suggests that the data is likely linearly separable or close to it. The Gaussian Kernel also performs well, indicating the presence of non-linear relationships that can be effectively captured with this kernel. The Polynomial kernels with higher degrees and the Sigmoid Kernel show lower performance, likely due to overfitting or the nature of the data.

By choosing the appropriate kernel, we can build robust and accurate models for breast cancer diagnosis, ultimately aiding in early detection and improving patient outcomes.

Bibliography

[1] J. Mairal, J.P. Vert, *Machine learning with kernel methods*, Lecture Notes, 2018. https://ncatlab.org/nlab/files/MarailVertKernelMethods.pdf

[2] C. Campbell, *An introduction to kernel methods*, Studies in Fuzziness and Soft Computing, 2001. https://calla.rnet.missouri.edu/cheng_courses/com2014/Quadratic_programming_svmintro.pdf

[3] T. Hofmann, B. Schölkopf, A.J. Smola, *A review of kernel methods in machine learning*, Max-Planck-Institute Technical Report, 2006. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9d14bc317c453ab38bf9f854b2a5f756cff20629

[4] T. Hofmann, B. Schölkopf, A.J. Smola, *Kernel methods in machine learning*, Annals of Statistics, 2008. https://projecteuclid.org/journals/annals-of-statistics/volume-36/issue-3/Kernel-methods-in-machine-learning/10.1214/009053607000000677.pdf

[5] F. Jäkel, B. Schölkopf, F.A. Wichmann, *A tutorial on kernel methods for categorization*, Journal of Mathematical Psychology, 2007. http://matt.colorado.edu/teaching/categories/jsw7.pdf

[6] B. Haasdonk, H. Burkhardt, *Invariant kernel functions for pattern analysis and machine learning*, Machine Learning, 2007. https://link.springer.com/content/pdf/10.1007/s10994-007-5009-7.pdf

[7] N. Cristianini, B. Scholkopf, *Support vector machines and kernel methods: the new generation of learning machines*, AI Magazine, 2002. https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/download/1655/1553/