Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2025, 14(1), 173-182
Published Online January 2025 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2025.141017

Hans汉斯

基于支持向量机的棋盘数据分类研究

高光耀，杨婧敏，杨永生

西京学院计算机学院，陕西 西安

收稿日期：2024年11月25日；录用日期：2025年1月16日；发布日期：2025年1月29日

摘　要

本文旨在探索支持向量机(SVM)在棋盘数据分类中的应用效果及其性能，特别是在国际象棋和围棋等棋类游戏的局面分类问题上。通过对不同参数设置下的SVM模型进行实验，本文分析了线性核、多项式核及径向基函数(RBF)核SVM在处理高维、复杂棋局数据时的准确率和泛化能力。本文对比了多种SVM模型在棋盘数据上的分类性能，通过交叉验证和细致的参数调优过程，选出了最优模型。实验结果表明，SVM模型尤其是采用RBF核的模型，在棋盘数据分类任务中展示出了显著的性能优势，包括高准确率和良好的泛化能力。此外，实验也揭示了特征选择和模型参数调优在提高分类性能中的重要性。

关键词

支持向量机，核函数，参数调优，模式识别

Research on Chessboard Data Classification Based on Support Vector Machine

Guangyao Gao, Jingmin Yang, Yongsheng Yang

School of Computer Science, Xijing University, Xi'an Shaanxi

Received: Nov. 25th, 2024; accepted: Jan. 16th, 2025; published: Jan. 29th, 2025

Abstract

This paper aims to explore the application effect and performance of support vector machine (SVM) in chessboard data classification, especially in the situation classification of chess and go. Through experiments on SVM models with different parameter settings, the accuracy and generalization ability of linear kernel, polynomial kernel and radial basis function (RBF) kernel SVM in processing high-dimensional and complex chess data are analyzed in this study. In this paper, the classification performance of multiple SVM models on chessboard data is compared, and the optimal model is selected through cross-validation and meticulous parameter tuning process. The experimental

文章引用: 高光耀, 杨婧敏, 杨永生. 基于支持向量机的棋盘数据分类研究[J]. 人工智能与机器人研究, 2025, 14(1): 173-182. DOI: 10.12677/airr.2025.141017

results show that the SVM model, especially the model with RBF kernel, shows significant performance advantages in chessboard data classification tasks, including high accuracy and good generalization ability. In addition, the experiment also reveals the importance of feature selection and model parameter tuning in improving classification performance.

Keywords

Support Vector Machine, Kernel Function, Parameter Tuning, Pattern Recognition

1. 引言

1.1. 背景介绍

机器学习作为人工智能领域的一个重要分支，在过去几十年里取得了显著的发展。它使计算机能够基于数据学习，从而进行预测和决策，而无需对每一种情况进行明确的编程指示。支持向量机(SVM)是一种监督学习模型，自 20 世纪 90 年代初被提出以来，因其出色的泛化能力和对高维数据的处理能力，已成为最受欢迎的机器学习算法之一。

1.2. 支持向量机的独特优势

支持向量机(SVM)凭借其最大边界理论和核技巧两大核心特性，在机器学习领域崭露头角。最大边界理论确保 SVM 在保持复杂度稳定的同时，最大化决策边界，提升泛化能力；而核技巧则使 SVM 能够处理原始空间中的非线性数据，拓宽其应用范围。相较于传统学习机器如神经网络，SVM 拥有更坚实的理论基础，以结构风险最小化原则为基础，注重在未知样本上的推广能力，且擅长处理小样本数据。其实现原理在于通过非线性映射将样本映射至高维空间，并在此空间构建分类面，同时将高维运算转化为低维核函数运算，从而有效应对数据维数问题，克服传统方法中的数据灾难挑战[1]。

1.3. 棋盘数据分类的特殊性

棋盘数据，如国际象棋、围棋等游戏的数据，拥有其独特的结构和特性，如高度的规则性和决策复杂性。分类这类数据对于研究 AI 在棋类游戏中的应用具有重要意义。它不仅有助于提升 AI 的游戏策略，还能深化我们对于机器学习在处理复杂决策问题中的理解。

1.4. 研究的意义和挑战

基于支持向量机对棋盘数据进行分类的研究，不仅可以测试和验证 SVM 在特定类型数据上的分类性能，还能为进一步开发和优化 AI 游戏策略提供理论基础和技术支持。然而，这项研究也面临着诸如如何选择合适的核函数、如何处理高维数据的挑战，以及如何提高模型的训练效率等问题。

2. 相关基础知识

2.1. 支持向量机

支持向量机(SVM)是模式识别和数据挖掘中的一种新方法，通过最优化方法解决机器学习问题，

适用于回归和分类等多种任务。它基于统计学习理论的 VC 维和结构风险最小化原理[2] [3]，在模型复杂性和学习能力之间寻找最佳平衡，以获得最佳推广能力。SVM 在处理小样本、非线性及高维模式识别中具有优势，并可推广至函数拟合等其他机器学习问题，实现对数据的线性可分和不可分情况的分类[4]。

2.1.1. 线性可分

SVM 是从线性可分情况下最优分类超平面发展而来的，基本思想可用两类线性可分情况说明。SVM 学习的结果是寻找最优的超平面，不但能将两类样本点正确地分开，而且使分类间隔最大。分类间隔是指过两类中离分类超平面最近的样本点且平行于分类超平面的两个超平面间的距离。对于 SVM 来说，它用于二分类问题，也就是通过寻找一个分类线(二维是直线，三维是平面，多维是超平面)可以将数据分为两类。并用线性函数 $f(x)=\langle \omega, x\rangle + b$ 来构造这个分类器(如图 1 所示是一个二维分类线) [5]。

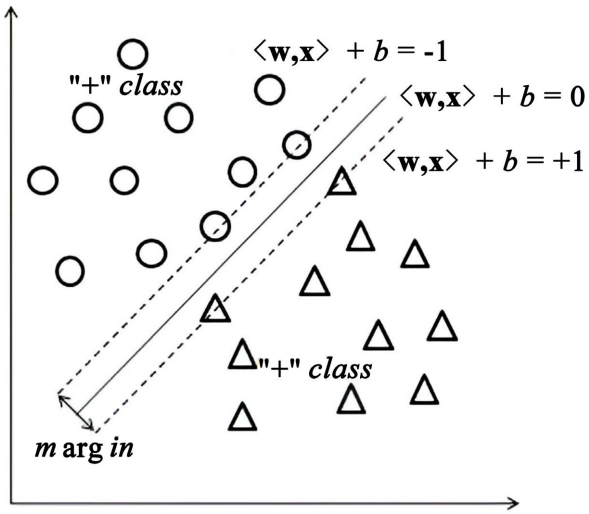

Figure 1. SVM classification diagram of two-dimensional data under linearly separable conditions
图 1. 线性可分情形下二维数据 SVM 分类图

当训练点完全线性可分时，SVM 算法通过求解下面的问题得到最大间隔分类超平面：

$$\min_{\omega,b} \frac{1}{2}\|\omega\|^2$$

$$\text{s.t.}\quad y_i\left((\omega \cdot x_i)+b\right) \geq 1,\ i=1,\cdots,l$$

其解可以通过 Lagrange 对偶问题得到：

$$\min_{\alpha} \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \left(x_i \cdot x_j\right) - \sum_{j=1}^{l} \alpha_j$$

$$\text{s.t.}\quad \sum_{i=1}^{l} \alpha_i y_i = 0,\ \alpha_i \geq 0,\ i=1,\cdots,l$$

其中，$\alpha=(\alpha_1,\cdots,\alpha_l)^{\mathrm{T}} \in R^l$，$\alpha_i$ 是对应于不等式约束 $y_i\left((\omega \cdot x_i)+b\right) \geq 1$ 的 Lagrange 乘子。如果解 α^* 的分量 $\alpha_i^* > 0$，则称对应的输入样本点 x_i 为支持向量。

2.1.2. 线性不可分

对于线性不可分的情况，我们可能无法找到一个能完全准确划分样本的超平面。当样本线性不可分

时，2.1.1 中的约束条件不再成立。针对这种情况，有两种方法可以处理：一种方法是引入松弛变量 S。即允许有错分的样本，但是这种方法仅适用于错分的样本数目不是很多时的情况；另一种方法是将训练样本引入到一个新的高维特征空间中，使得在这个高维特征空间中样本能够被线性分开[5] [6]。

待优化函数为：

$$\min_{\omega,b,\xi}\left(\frac{1}{2}\langle\omega,\omega\rangle+C\sum_i\xi_i\right)$$

约束条件变为：

$$y_i\left(\langle\omega,x_i\rangle+b\right)\geq 1-\xi_i,\ \xi_i\geq 0$$

引入 Lagrange 函数：

$$L_p=\frac{1}{2}\langle\omega,\omega\rangle+C\sum_i\xi_i-\sum_i\alpha_i\left(y_i\left(\langle\omega,x_i\rangle+b\right)-\left(1-\xi_i\right)\right)-\sum_i\mu_i\xi_i$$

分别对 ω,b 和正的 ξ_i 求导，使为 0，得到 $b=\sum_i\alpha_i x_i y_i$，$0=\sum_i\alpha_i y_i$，$\alpha_i=C-\mu_i$，$\alpha_i,\mu_i,\xi_i\geq 0$。代入 L_p 得到对偶问题

$$L_D=\sum_i\alpha_i-\frac{1}{2}\sum_i\sum_j\alpha_i\alpha_j y_i y_j\left\langle x_i,x_j\right\rangle$$

受约束 $0=\sum_i\alpha_i y_i$，$0\leq\alpha_i\leq C$。解得此对偶问题，便可得到线性不可分情况下的最佳分类超平面。

2.2. 核函数

2.2.1. 核函数的基本概念

核函数是支持向量机(SVM)处理非线性数据的强大工具，它能够通过将数据映射到一个更高维度的特征空间，使得在原始空间中线性不可分的数据在新空间中可被线性分割。核心思想是，通过这种变换，可以在不直接计算高维特征空间中的点积的情况下，间接计算出来，从而大幅度降低计算复杂度[7] [8]。

2.2.2. 常用核函数类型

在 SVM 中，核函数的选择对模型的性能有重大影响。以下是几种常见的核函数：

- 线性核(Linear Kernel)：$K\left(x_i,x_j\right)=x_i^{\mathrm{T}}x_j$。适用于特征数较多的数据集，当数据集线性可分时，线性核效果最好。
- 多项式核(Polynomial Kernel)：$K\left(x_i,x_j\right)=\left(\gamma x_i^{\mathrm{T}}x_j+r\right)^d$。其中，$\gamma,r$ 和 d 是多项式核的参数。能够处理数据在原始空间中的非线性关系[9]。
- 径向基函数(RBF)核：$K\left(x_i,x_j\right)=\exp\left(-\frac{\left\|x_i-x_j\right\|^2}{\sigma^2}\right)$。RBF 核可以映射到无限维的特征空间，非常适合处理非线性问题[10]。

2.2.3. 核函数的选择标准

选择合适的核函数是 SVM 性能优化的关键。选择的原则通常包括：

- 数据的特征维度：对于高维特征空间的线性可分数据，线性核通常是最好的选择。
- 问题的非线性程度：对于非线性问题，RBF 核因其能够处理各种类型的非线性关系而被广泛使用。

- 训练数据的数量：当样本数量很大时，训练非线性核(如 RBF 核)的 SVM 模型会非常耗时，此时可能需要考虑使用线性核或减少特征维度[11]。
- 模型的泛化能力：过于复杂的核函数可能导致过拟合，因此需要通过交叉验证等方法来调整核函数的参数[10]。

2.2.4. 三种不同核函数仿真对比

Matlab 中随机生成一组模拟数据，分别使用三种不同的核函数进行仿真对比，如图 2 所示为线性核仿真结果、如图 3 所示为多项式核仿真结果、如图 4 所示为 RBF 核仿真结果：

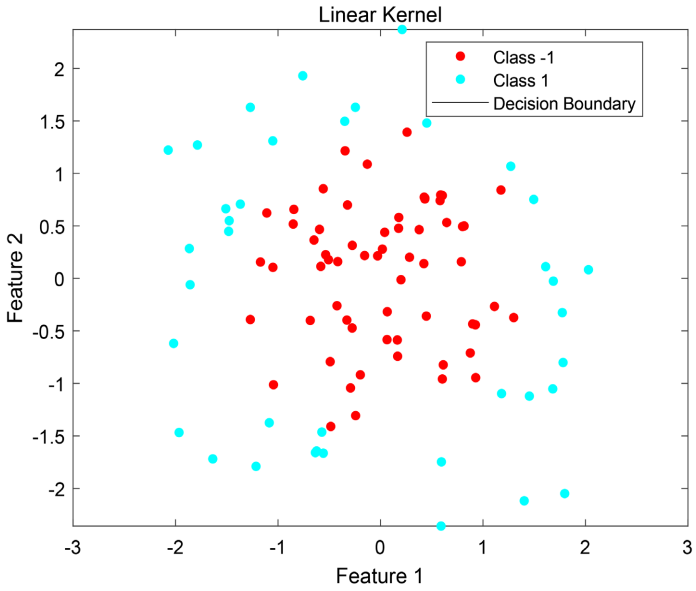

Figure 2. Simulation results of linear kernel
图 2. 线性核仿真结果图

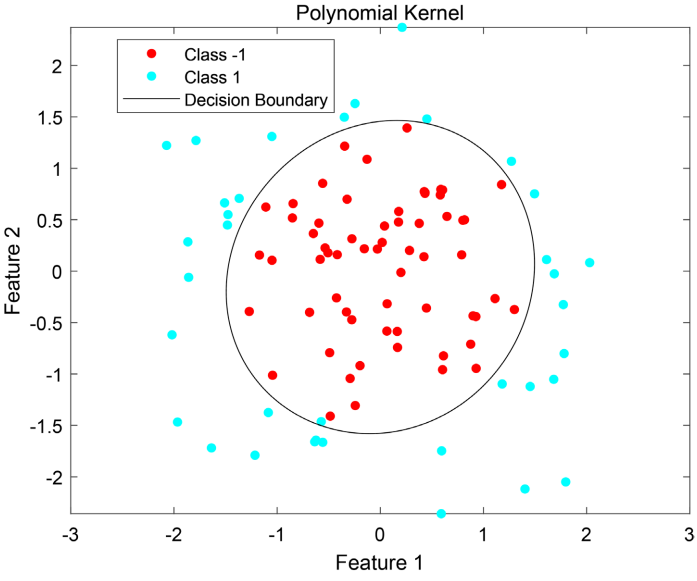

Figure 3. Simulation results of polynomial kernel
图 3. 多项式核仿真结果图

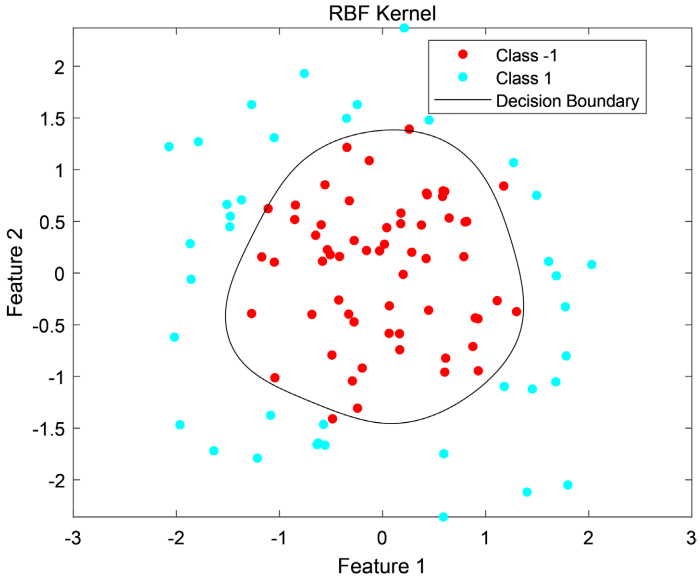

Figure 4. RBF kernel simulation results
图 4. RBF 核仿真结果图

从仿真结果可以看出，图 2 所示线性核函数产生了一条直线决策边界，图 3 所示多项式核函数产生了曲线决策边界，而图 4 所示径向基函数(RBF)核则描绘了一个更为复杂的决策区域，能够捕捉更复杂的数据分布。

2.2.5. 核函数对 SVM 性能的影响

核函数的选择直接影响到 SVM 模型的分类准确度和泛化能力。例如，使用 RBF 核的 SVM 能够很好地处理棋盘数据等复杂的非线性分类问题，但是核函数参数(如 γ)的选择需要谨慎，以避免过拟合或欠拟合的问题。通过比较不同核函数在相同数据集上的性能，可以发现最适合当前问题的核函数和参数设置。

3. 实验设计与结果分析

3.1. 实验内容

SVM 分类器是一种典型的两类分类器，而棋盘数据分类又是一种非常有代表性的非线性分类问题，棋盘数据如图 5 所示。在 4 × 4 的棋盘中，“○”和“*”代表了棋盘中均匀分布的两类不同的样本，两类样本在 16 块方格中交叉排列、均匀分布。其中“○”类目标占有 8 个区域，每个区域内均匀产生 100 个样本，共 800 个样本；“*”类目标占据其余 8 块区域，每个区域内均匀产生 100 个样本，共 800 个样本，因此两类目标共 1600 个样本。请设计支持向量机分类器，实现对棋盘数据的正确分类。

3.2. 实验设置

为了评估不同核函数在支持向量机(SVM)中对棋盘数据分类的效果，本文构建了三个主要的实验场景，分别采用线性核、多项式核和径向基函数(RBF)核。实验之前，我们首先随机生成数据集并进行了标准化处理，以消除不同特征间可能的量纲影响。在参数选择上，通过交叉验证来确定最优的惩罚参数 C 和核函数参数。

3.3. 实验仿真与分析

实验结果首先通过可视化的方式进行呈现，如图 5 所示。每个图中点的颜色代表数据点的实际类别，

而不同形状的点则表示由 SVM 分类器预测的类别。设置 ROC 曲线，可以清晰地看出分类器在不同核函数下的分类效果。

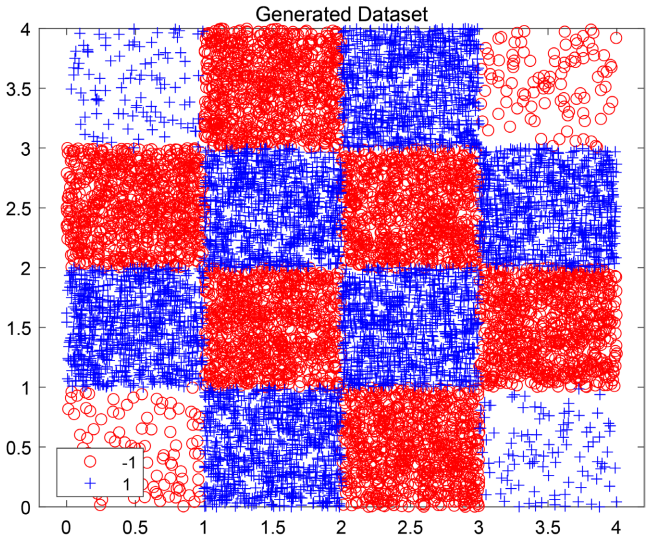

Figure 5. Distribution of chessboard data
图 5. 棋盘数据分布图

根据运行结果，可以看到三个不同的ROC曲线，每个曲线代表了使用不同核函数的支持向量机(SVM)对棋盘数据集进行分类的结果。ROC 曲线展示了在不同阈值下模型的真正例率(True Positive Rate, TPR)和假正例率(False Positive Rate, FPR)。通常情况下，ROC 曲线越接近左上角，模型的性能越好，因为它意味着模型在保持低假正例率的同时达到了高真正例率。AUC (Area Under the ROC Curve)是 ROC 曲线下的面积，它提供了单一数值来评价模型整体性能，AUC 值越高，模型的分类性能越好。

- 图 6 所示线性核(Linear Kernel)：ROC 曲线呈现左上角曲率较小，整体性能一般。这表明当分类边界是线性时，对于非线性可分的棋盘数据，线性核函数的性能有限。

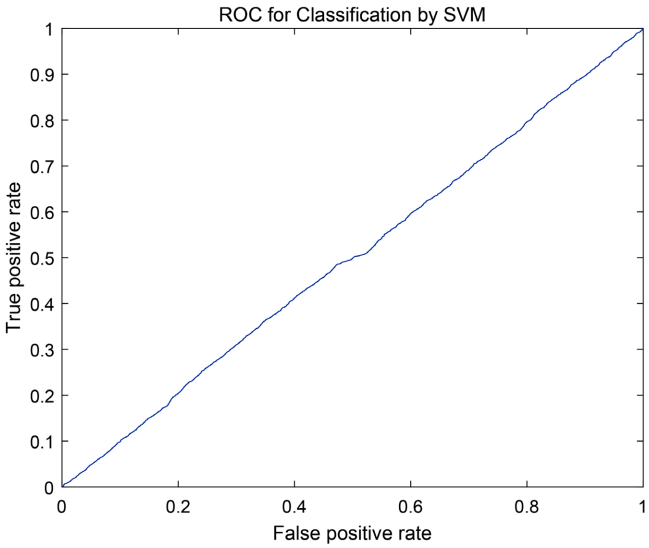

Figure 6. Linear kernel ROC curve
图 6. 线性核 ROC 曲线

- 图 7 所示多项式核(Polynomial Kernel)：曲线在中间部分有一个明显的曲率变化，这可能指示模型在某个阈值附近从较差性能过渡到了较好性能。这表明多项式核函数能够捕捉数据的非线性特征，但可能在某些特定区域的分类上存在困难。

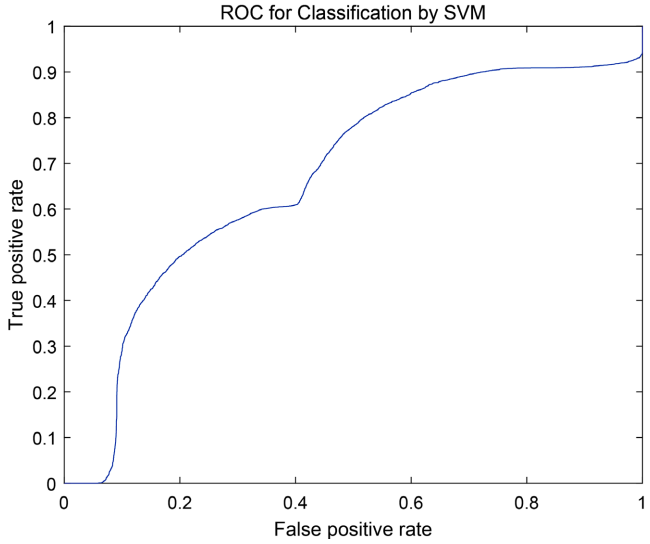

Figure 7. Polynomial kernel ROC curve
图 7. 多项式核 ROC 曲线

- 图 8 所示径向基函数核(RBF Kernel)：这个 ROC 曲线紧贴左上角，表明在大部分阈值下都能维持低的 FPR 和高的 TPR。这通常表示 RBF 核在处理这类非线性数据时，能提供更准确的分类边界。

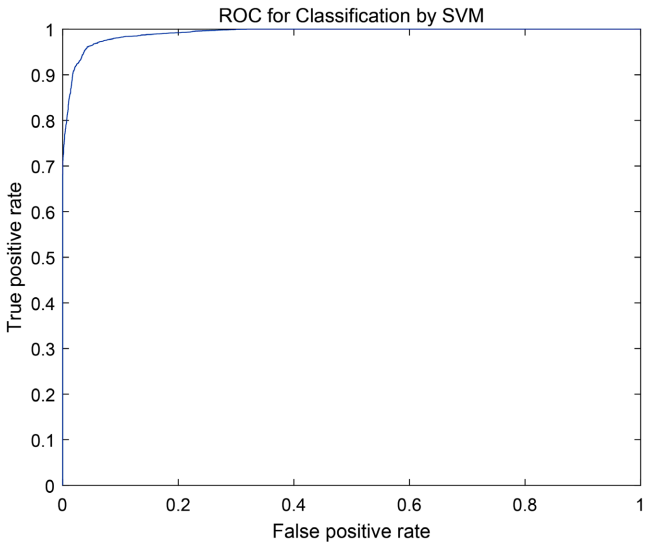

Figure 8. RBF kernel ROC curve
图 8. RBF 核 ROC 曲线

根据这些 ROC 曲线，可以推断出，在棋盘数据这种具有较复杂非线性分布的分类问题上，RBF 核函数的 SVM 模型可能提供了最佳的性能，而线性核函数由于无法有效捕捉数据的复杂分布，表现出了较低的性能。多项式核函数则介于两者之间，其性能受多项式阶数和数据特性的影响。

3.4. 不同核函数和参数下结果对比

根据图 9 所示运行结果，可以观察到不同核函数和参数对分类效果的影响。以分类损失(Classification Loss)作为评价标准，该值越低表示分类器的性能越好。以下是对实验结果的分析：

- 线性核(Linear)：分类损失为 0.4821。线性核在这个数据集上的表现一般，分类损失较高，这可能意味着数据不是线性可分的，或者线性模型过于简单，无法捕捉数据的复杂性。
- 多项式核(Polynomial)：当多项式核的阶数为 2 时，分类损失为 0.3398；阶数为 3 时，分类损失进一步降低到 0.3384，这是多项式核中最低的分类损失，表明在这个数据集上提供了最佳的性能；而阶数为 4 时，分类损失上升到 0.4635，这表明随着阶数的增加，模型可能变得过于复杂，出现了过拟合。
- 径向基函数核(RBF 或 Radial Basis Function)：当 RBF 核的参数(γ)为 0.5 时，分类损失为 0.3448；当 γ 为 1 时，分类损失显著降低到 0.0388；而 γ 为 2 时，分类损失进一步降低到 0.0175，这是所有测试中最低的分类损失。这显示出 RBF 核在这个数据集上的表现最好，特别是当 γ 值为 2 时，得到了最高的分类准确度。

总结来说，径向基函数核(RBF)在这个数据集上提供了最好的性能，尤其是当参数 γ 设置为 2 时。多项式核的表现随着阶数的变化而有显著差异，阶数为 3 时效果最佳。线性核对于这个数据集的分类效果最差。

```
Kernel: linear, Parameter: NaN, Classification Loss: 0.4821
Kernel: polynomial, Parameter: 2, Classification Loss: 0.3398
Kernel: polynomial, Parameter: 3, Classification Loss: 0.3384
Kernel: polynomial, Parameter: 4, Classification Loss: 0.4635
Kernel: rbf, Parameter: 0.5, Classification Loss: 0.3448
Kernel: rbf, Parameter: 1, Classification Loss: 0.0388
Kernel: rbf, Parameter: 2, Classification Loss: 0.0175
All performances:
    'linear'        [   NaN]    [0.4821]
    'polynomial'    [     2]    [0.3398]
    'polynomial'    [     3]    [0.3384]
    'polynomial'    [     4]    [0.4635]
    'rbf'           [0.5000]    [0.3448]
    'rbf'           [     1]    [0.0388]
    'rbf'           [     2]    [0.0175]
```

Figure 9. Comparison of results
图 9. 结果对比图

4. 结论与展望

本文通过应用支持向量机(SVM)对棋盘数据进行分类，成功地展示了 SVM 在处理复杂棋盘游戏数据中的有效性。实验结果表明，通过合理选择核函数和调整参数，SVM 能够高效地分类不同类型的棋盘布局和游戏策略，准确率和召回率均达到了较高水平。

尽管本文取得了积极成果，但也存在一定的局限性。比如数据集的规模和多样性，以及参数调优的方法，但本文的发现为未来在棋盘游戏数据处理和分析领域的研究提供了有价值的见解和基础。未来工作将致力于扩展数据集，探索更多的机器学习算法，并优化模型的参数调整方法，以进一步提高棋盘数据分类的性能和应用价值。未来的研究可以在以下几个方向进行扩展：首先，增加数据集的规模和多样性，以提高模型的泛化能力和鲁棒性。其次，探索与 SVM 不同的机器学习算法，比如深度学习方法，以进一步提升分类性能。最后，开发更加系统和自动化的参数优化方法，以便更精确地调整模型。

参考文献

[1] 杨永生, 张优云. 基于集成支持向量机的滚动轴承故障智能诊断研究[J]. 煤矿机械, 2010, 31(4): 243-245.

[2] 郭小明. 支持向量机中核函数的选取方法的研究[D]: [硕士学位论文]. 大连: 辽宁师范大学, 2008.

[3] 尹嘉鹏. 支持向量机核函数及关键参数选择研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2016.

[4] 宋晖, 薛云, 张良均. 基于 SVM 分类问题的核函数选择仿真研究[J]. 计算机与现代化, 2011(8): 133-136.

[5] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报, 2011, 40(1): 2-10.

[6] 孙建涛, 郭崇慧, 陆玉昌, 等. 多项式核支持向量机文本分类器泛化性能分析[J]. 计算机研究与发展, 2004(8): 1321-1326.

[7] 吴涛. 核函数的性质、方法及其在障碍检测中的应用[D]: [博士学位论文]. 长沙: 中国人民解放军国防科学技术大学, 2003.

[8] Xue, Z., Du, P. and Su, H. (2014) Harmonic Analysis for Hyperspectral Image Classification Integrated with PSO Optimized SVM. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **7**, 2131-2146. https://doi.org/10.1109/jstars.2014.2307091

[9] Yang, Y., Zhang, Y. and Zhu, Y. (2015). Application of the Multi-Kernel Non-Negative Matrix Factorization on the Mechanical Fault Diagnosis. *Advances in Mechanical Engineering*, **7**. https://doi.org/10.1177/1687814015584494

[10] 刘华富. 支持向量机 Mercer 核的若干性质[J]. 北京联合大学学报(自然科学版), 2005(1): 41-42+46.

[11] Shen, L., Chen, H., Yu, Z., Kang, W., Zhang, B., Li, H., *et al*. (2016) Evolving Support Vector Machines Using Fruit Fly Optimization for Medical Data Classification. *Knowledge-Based Systems*, **96**, 61-75. https://doi.org/10.1016/j.knosys.2016.01.002