

基于弹球损失模糊支持向量机的客户认购定期存款预测分析

匡维波

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2024年6月26日; 录用日期: 2024年7月8日; 发布日期: 2024年8月22日

摘要

本文基于UCI机器学习库中的一家银行机构营销活动的的数据, 以客户是否认购定期存款为响应变量, 以9个描述客户信息以及社会经济状况的离散和连续指标作为解释变量建立预测模型。提出了一种改进的弹球损失模糊支持向量机(Pin-FSVM)预测模型, 旨在提高金融服务行业中银行识别潜在客户认购定期存款的准确性和效率。Pin-FSVM模型通过融合弹球损失函数和模糊隶属度的概念, 优化了传统模糊支持向量机的性能。该模型在含噪声数据环境下维持了预测准确率, 并有效处理了数据中的不确定性。在应用于银行客户认购存款的预测实践中, Pin-FSVM成功识别出更可能选择定期存款的客户, 显著提升了预测效果, 为银行精准识别和服务客户群体提供了有力工具。

关键词

弹球损失函数, 模糊支持向量机, 定期存款, 预测模型

Forecasting Analysis of Customer Subscribed Time Deposits Based on Pinball Loss Fuzzy Support Vector Machines

Weibo Kuang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Jun. 26th, 2024; accepted: Jul. 8th, 2024; published: Aug. 22nd, 2024

Abstract

This paper is based on data from the UCI machine learning repository on the marketing activities

文章引用: 匡维波. 基于弹球损失模糊支持向量机的客户认购定期存款预测分析[J]. 电子商务评论, 2024, 13(3): 8568-8578. DOI: 10.12677/ecl.2024.1331049

of a banking institution, with whether a customer subscribes to a time deposit as the response variable, and nine discrete and continuous indicators describing the customer's information as well as his socio-economic status as the explanatory variables. In this paper, an improved Pinball Loss Fuzzy Support Vector Machine (Pin-FSVM) prediction model is proposed with the aim of improving the accuracy and efficiency of banks in the financial services industry in identifying potential customers to subscribe to time deposits. The Pin-FSVM model optimises the performance of the traditional fuzzy support vector machine by incorporating the concepts of the Pinball Loss Function and the Fuzzy Affiliation Degree. The model maintains prediction accuracy in noisy data environments and effectively handles the uncertainty in the data. In the application to the practice of predicting bank customers' subscription deposits, Pin-FSVM successfully identifies customers who are more likely to choose time deposits, which significantly improves the prediction effect and provides a powerful tool for banks to accurately identify and serve their customer groups.

Keywords

Pinball Loss Function, Fuzzy Support Vector Machines, Time Deposits, Predictive Models

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

定期存款亦称“定期存单”，银行与存款人双方在存款时事先约定期限、利率，到期后支取本息的存款。有的定期存单到期时，储户所需款项可上市出售；有的定期存单无法转让，存款人若选择到期前从银行取款，则需从银行缴纳一定费用。伴随着中国经济发展进入新常态、金融监管不断加强、行业经营竞争不断加剧，再加上互联网金融的影响，对中国商业银行整体经营环境造成很大的影响[1]。随着更多外资银行大量进入中国市场，各家银行之间对优质客户的争夺以及对市场份额的拓展也愈演愈烈，各家银行该如何赢得竞争，依赖于能够从海量的顾客数据中发掘出潜在的宝贵信息，增强顾客事务处理能力，深层次地发掘顾客和合理定位顾客，从而达到高效管理的目的。顾客是商业银行争夺的重点，银行利润归根到底来自于顾客，为了增强盈利能力就必须重视对顾客价值的挖掘。从顾客需求角度看，不同种类的顾客需求存在差异，要使不同顾客对同一家银行均满意，需要银行有针对性地提供满足顾客需要的产品与服务，而要想满足这一多样化异质需求，必须根据不同标准细分客户群体；在顾客价值层面上，不同顾客能给银行带来的价值也不一样，银行需要了解什么顾客对银行有高价值，什么顾客对银行忠诚，银行潜在客户有哪些、成长性最强、最易流失的客户有哪些，银行必须将其客户细分；从银行资源与能力角度出发，如何优化配置不同客户的有限资源是每一家银行必须思考的问题，而商业银行的服务资源是有限的，顾客的数量是很多的，因此在顾客管理中十分需要进行顾客的统计与分析。如何建立合理的模型预测客户是否认购定期存款十分重要。

Vapnik 和 Cortes 于 20 世纪末提出的一种全新的分类方法[2]：支持向量机(Support vector machine, SVM)。随着 SVM 理论与算法日趋成熟，其广泛应用于多个领域如模式识别、心理分析及医学影像检测，更多相关应用场景可以参考相关文献[3]-[5]。Huang 等人将弹球损失函数应用于 SVM，提出了基于弹球损失函数的支持向量机(Pin-SVM) [6]。同时研究了 Pin-SVM 的性质，包括噪声不敏感性，鲁棒性和误分类误差最小化。与经典的 SVM 相比，Pin-SVM 具有相同的计算复杂度，并且具有噪声不敏感性和重采样稳定性。为了进一步提升 SVM 处理离群点的性能，Lin 等人[7]提出了模糊支持向量机(Fuzzy support

vector machine, FSVM), 通过引入模糊隶属度以表征不同样本的重要程度, 为不同样本设置不同的模糊隶属度以反映其对分类的重要性, 在一定程度上降低了异常值对决策函数的影响, 有效缓解了 SVM 对异常值敏感的问题。计算模糊隶属度值的策略对于 FSVM 方法的性能至关重要, 传统 FSVM 算法在设计模糊隶属度时常以距离作为衡量样本重要性的指标。Jiang 等人[8]将隶属度的计算进行核扩展, 在高维特征空间中判定隶属度大小, 一定程度上减小了隶属度函数对样本集几何结构的依赖性, 使模型对于异常点或噪声的判定更加准确。

受 FSVM, Pin-SVM 的启发, 本文提出了一种新的基于弹球损失的模糊支持向量机(Pin-FSVM)模型应用于对客户是否认购定期存款的预测模型, 提高预测精度。主要思想是将 FSVM 与 Pin-SVM 结合, 将 FSVM 中的铰链损失函数替换为弹球损失函数, 减少噪声对决策超平面的影响, 使 Pin-FSVM 在处理噪声数据时具有稳健性。Pin-FSVM 使用松弛变量的 L_1 范数, 所以提出的 Pin-FSVM 是凸二次规划问题, 可以使用拉格朗日乘子法求解对偶问题, 然后通过 CVXR (求解约束凸优化问题的优化包)求解模型。改进的基于弹球损失的模糊支持向量机对异常值具有鲁棒性, 为银行更加精确地识别客户是否认购定期存款提供了方案。

2. 预备知识

在本章中, 对 SVM、Pin-SVM、FSVM 的理论进行简要的介绍, 在一个二分类问题中, 假设输入空间中训练样本集为 $\{x_i, y_i\}_{i=1}^n$, 其中 $x_i \in \mathbb{R}^m$, $y_i \in \{-1, 1\}$ 。

2.1. 基于弹球损失函数的支持向量机

弹球损失在回归分析中得到了深入的研究和广泛的应用, 但尚未用于分类, Huang 等人为了克服 SVM 对决策超平面附近的特征噪声敏感提出了基于弹球损失函数的支持向量机。首先简单介绍弹球损失 (Pinball loss) 函数:

$$L_{\tau}(x, y, f(x)) = \begin{cases} 1 - yf(x), & 1 - yf(x) \geq 0, \\ -\tau(1 - yf(x)), & 1 - yf(x) < 0. \end{cases} \quad (1)$$

其中, $0 \leq \tau \leq 1$ 为自定义参数, 图 1 为参数 $\tau = 0.1, 0.2, 0.5, 1$ 时的函数图像。

基于弹球损失函数的支持向量机(Pin-SVM)模型定义如下:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i, \\ & y_i (w^T x_i + b) \leq 1 + \frac{1}{\tau} \xi_i, i = 1, 2, \dots, n. \end{aligned} \quad (2)$$

通过拉格朗日乘子法可求得(2)的对偶问题为:

$$\begin{aligned} \min_{\alpha, \beta} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \beta_i) (\alpha_j - \beta_j) x_i^T x_j y_i y_j - \sum_{i=1}^n (\alpha_i - \beta_i) \\ \text{s.t.} \quad & \sum_{i=1}^n (\alpha_i - \beta_i) y_i = 0, \\ & \alpha_i + \frac{1}{\tau} \beta_i = C, \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (3)$$

其中 $\alpha_i \geq 0, \beta_i \geq 0, i = 1, 2, \dots, n$ 为拉格朗日乘子, 线性情况下 Pin-SVM 决策函数为:

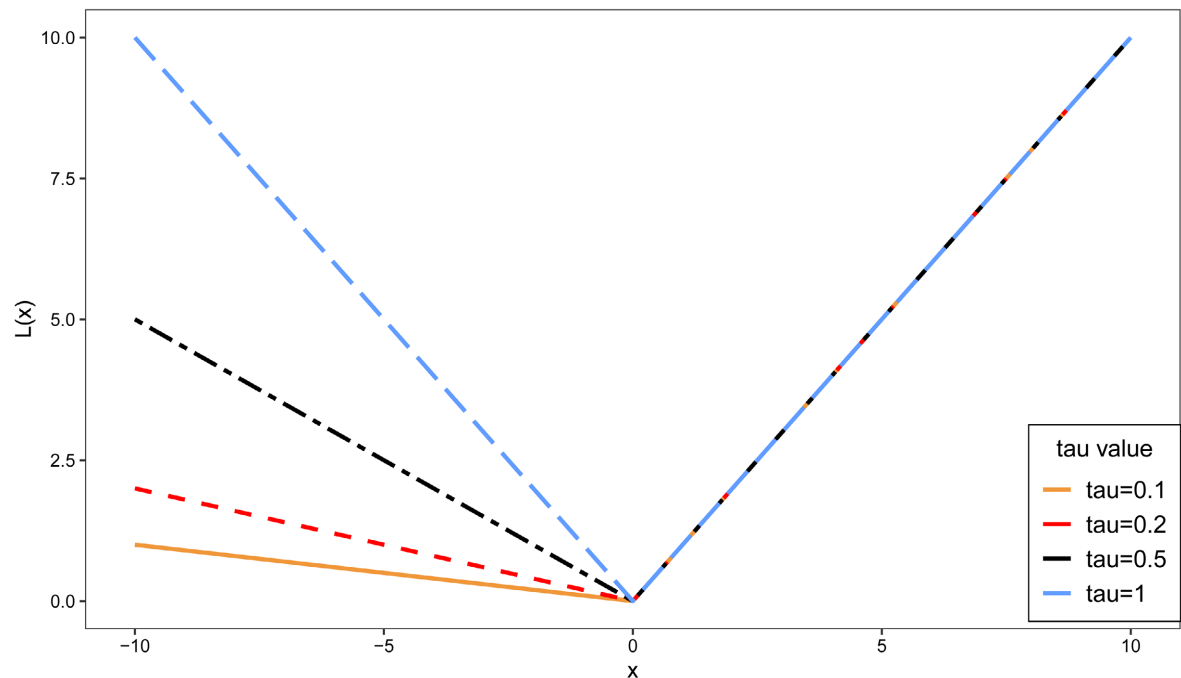


Figure 1. Image of the pinball loss function when τ taking different values

图 1. 不同的 τ 取值时弹球损失函数图像

$$f(x) = \text{sign}\left(\sum_{i=1}^n (\alpha_i - \beta_i) y_i x^T x_i + b\right). \quad (4)$$

随着参数 τ 的值增加, 正确分类的样本点的权重变大, 因此边距宽度变大。所以类边界附近的点在决定最优决策超平面中变得不那么重要。对于 Pin-SVM, 在正确分类的样本点上给予惩罚带来了决策边界周围的噪声的不敏感性, 特征噪声的影响被削弱, 使得 Pin-SVM 对特征噪声不敏感。

2.2. 基于 Hinge 损失的模糊支持向量机

在传统的 SVM 中, 输入数据的每个样本点的重要程度相同, 但是对于许多现实生活的分类问题, 训练分类模型时某些样本点可能比其它样本点更重要。所以提出了 FSVM 来解决这个问题, FSVM 中模糊隶属函数根据每个样本点的重要性分配模糊隶属值, 使用类中心方法来生成模糊隶属度。线性情形下, 令 x_+ 和 r_+ 分别表示 +1 类的均值和半径, x_- 和 r_- 表示 -1 类的均值和半径。其中 $r_+ = \max_{x_i, y_i=1} \|x_+ - x_i\|$, $r_- = \max_{x_i, y_i=-1} \|x_- - x_i\|$ 。对于训练样本点 x_i , 其模糊隶属度 s_i 定义如下:

$$s_i = \begin{cases} 1 - \|x_+ - x_i\| / (r_+ + d), & \text{if } y_i = +1, \\ 1 - \|x_- - x_i\| / (r_- + d), & \text{if } y_i = -1. \end{cases} \quad (5)$$

为了避免模糊隶属度 s_i 为 0, 设置常数 $d > 0$ 。

对于非线性分类问题, 使用改进的核模糊隶属函数, 定义为:

$$s_i = \begin{cases} 1 - \sqrt{\|d_i^+\|^2 / ((r_+)^2 + d)}, & \text{if } y_i = +1, \\ 1 - \sqrt{\|d_i^-\|^2 / ((r_-)^2 + d)}, & \text{if } y_i = -1. \end{cases} \quad (6)$$

式(15)中,

$$\begin{aligned} (d_i^+)^2 &= k(x_i, x_i) - \frac{2}{m^+} \sum_{(x_j, y_j=+1)} k(x_i, x_j) + \frac{2}{(m^+)^2} \sum_{(x_j, y_j=+1, x_k, y_k=+1)} k(x_j, x_k), \\ (r^+)^2 &= \max_{x, y=+1} \left\{ k(x_i, x_i) - \frac{2}{m^+} \sum_{(x_j, y_j=+1)} k(x_i, x_j) + \frac{2}{(m^+)^2} \sum_{(x_j, y_j=+1, x_k, y_k=+1)} k(x_j, x_k) \right\}, \\ (d_i^-)^2 &= k(x_i, x_i) - \frac{2}{m^-} \sum_{(x_j, y_j=-1)} k(x_i, x_j) + \frac{2}{(m^-)^2} \sum_{(x_j, y_j=-1, x_k, y_k=-1)} k(x_j, x_k), \\ (r^-)^2 &= \max_{x, y=-1} \left\{ k(x_i, x_i) - \frac{2}{m^-} \sum_{(x_j, y_j=-1)} k(x_i, x_j) + \frac{2}{(m^-)^2} \sum_{(x_j, y_j=-1, x_k, y_k=-1)} k(x_j, x_k) \right\}. \end{aligned}$$

其中 m^+ 和 m^- 分别代表正类样本数和负类样本数。样本与模糊隶属度为 $(x_1, y_1, s_1), (x_2, y_2, s_2), \dots, (x_n, y_n, s_n)$, 提出的模糊支持向量机定义为:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n s_i \xi_i \\ \text{s.t.} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (7)$$

其中 C 是常数, 较小的 s_i 减少了问题(7)中的参数 ξ_i 的影响, 使得其相对应的样本点被视为不太重要的样本点。

3. 基于弹球损失的模糊支持向量机(Pin-FSVM)

3.1. 线性情况

结合 Pin-SVM 和 FSVM, 本节提出了基于弹球损失函数的模糊支持向量机(Pin-FSVM), 将 FSVM 中的较链损失函数替换为弹球损失函数, 增强 FSVM 的性能。使用松弛变量的 L_1 范数, 所以提出的 Pin-FSVM 是凸二次规划问题, 可以通过求解经典的凸二次规划问题, 利用对偶问题来求解模型, 其时间复杂化度与 FSVM 相同。

将弹球损失函数引入到传统的 FSVM 中, 可得线性情况下 Pin-FSVM 的目标函数:

$$\min_{w, b, \xi} \quad \frac{1}{2} w^T w + C \sum_{i=1}^n s_i L_\tau (1 - y_i (w^T x_i + b)). \quad (8)$$

其等价约束形式如式(9)所示:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n s_i \xi_i \\ \text{s.t.} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i, \\ & y_i (w^T x_i + b) \leq 1 + \frac{1}{\tau} \xi_i, i = 1, 2, \dots, n. \end{aligned} \quad (9)$$

其中 $C > 0$ 和 $0 \leq \tau \leq 1$ 为自定义参数, s_i 和 ξ_i 分别为第 i 个输入训练样本点的模糊隶属度和松弛变量。

求解的式(9)是一个经典的凸二次规划问题, 通过求解其对偶问题的解得到原问题的解, 引入拉格朗日乘子法可得其拉格朗日函数, 首先求解线性情况下的 Pin-FSVM 对偶问题, 拉格朗日函数为:

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} w^T w + C \sum_{i=1}^n s_i \xi_i - \sum_{i=1}^n \alpha_i \left(y_i (w^T x_i + b) - 1 + \xi_i \right) - \sum_{i=1}^n \alpha_i \left(-y_i (w^T x_i + b) + 1 + \frac{1}{\tau} \xi_i \right), \quad (10)$$

其中 $\alpha_i > 0, \beta_i > 0, i = 1, 2, \dots, n$ 为拉格朗日乘子, 在 $L(w, b, \xi, \alpha, \beta)$ 中对 w, b, ξ_i 求偏导, 并令其偏导数为 0, 可得:

$$\begin{cases} \frac{\partial L}{\partial w} = w - \sum_{i=1}^n (\alpha_i - \beta_i) y_i x_i = 0, \\ \frac{\partial L}{\partial b} = -\sum_{i=1}^n (\alpha_i - \beta_i) y_i = 0, \\ \frac{\partial L}{\partial \xi_i} = s_i C - \alpha_i - \frac{1}{\tau} \beta_i = 0, \quad \forall i = 1, 2, \dots, n. \end{cases} \quad (11)$$

所以式(9)的对偶问题转化为:

$$\begin{aligned} \min_{\alpha, \beta} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \beta_i) (\alpha_j - \beta_j) x_i^T x_j y_i y_j - \sum_{i=1}^n (\alpha_i - \beta_i) \\ \text{s.t.} \quad & \sum_{i=1}^n (\alpha_i - \beta_i) y_i = 0, \\ & \alpha_i + \frac{1}{\tau} \beta_i = s_i C, \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (12)$$

其中, $\alpha_i \geq 0, \beta_i \geq 0, i = 1, 2, \dots, n$ 是拉格朗日乘子。

为了方便在求解器中求解所提出模型的参数, 将 Pin-FSVM 模型的对偶问题转换为矩阵形式, 其中 $e = [1, 1, \dots, 1]_{n \times 1}^T$, E 为 $n \times n$ 的单位矩阵, $D = \text{diag}(y_1, y_2, \dots, y_n)$, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]_{n \times 1}^T$, $\beta = [\beta_1, \beta_2, \dots, \beta_n]_{n \times 1}^T$,

$$s = [s_1, s_2, \dots, s_n]_{n \times 1}^T, \quad u = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}_{2n \times 1}, \quad X \text{ 为样本矩阵, } X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}_{n \times m},$$

$$D_{mat} = \begin{bmatrix} DXX^T D & -DXX^T D \\ -DXX^T D & DXX^T D \end{bmatrix}_{2n \times 2n}, \quad G = \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix}_{2n \times 2n}, \quad \text{可将其对偶问题(12)的转化为以下矩阵形式:}$$

$$\begin{aligned} \min_u \quad & \frac{1}{2} u^T D_{mat} u - [e^T, -e^T] u \\ \text{s.t.} \quad & [e^T, -e^T] G u = 0, \\ & \left[E, \frac{1}{\tau} E \right] u = sC, u \geq 0. \end{aligned} \quad (13)$$

在线性情况下通过求解对偶问题(13), 可求得 $w = \sum_{i=1}^n (\alpha_i - \beta_i) y_i x_i$, 当 $0 < \alpha_i - \beta_i < s_i C$ 时, 对应的样本为支持向量, 任意的支持向量 (x_s, y_s) 均满足 $y_s f(x_s) = 1$, 即表示为:

$$y_s \left(\sum_{i \in S} (\alpha_i - \beta_i) y_i x_i^T x_s + b \right) = 1, \quad (14)$$

其中 $S = \{i | \alpha_i > 0, i = 1, 2, \dots, m\}$ 为所有支持向量的下标集, 使所有支持向量求解的平均值求解偏置项:

$$b = \frac{1}{|S|} \sum_{i \in S} \left(\frac{1}{y_i} - \sum_{i \in S} \alpha_i y_i x_i^T x_s \right), \quad (15)$$

求解出参数 w, b 后即可得 Pin-FSVM 决策函数为:

$$f(x) = \text{sign} \left(\sum_{i=1}^n (\alpha_i - \beta_i) y_i x^T x_i + b \right). \quad (16)$$

3.2. 非线性情况

与基于铰链损失函数的模糊支持向量机类似, 通过引入非线性映射 $\phi(x)$, Pin-FSVM 在非线性情况下目标函数同理得出:

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^n s_i L_r \left(1 - y_i (w^T \phi(x_i) + b) \right). \quad (17)$$

其等价约束形式如(18)所示:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n s_i \xi_i \\ \text{s.t.} \quad & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & y_i (w^T \phi(x_i) + b) \leq 1 + \frac{1}{\tau} \xi_i, i = 1, 2, \dots, n. \end{aligned} \quad (18)$$

拉格朗日函数为:

$$\begin{aligned} L(w, b, \xi, \alpha, \beta) = & \frac{1}{2} w^T w + C \sum_{i=1}^n s_i \xi_i - \sum_{i=1}^n \alpha_i \left(y_i (w^T \phi(x_i) + b) - 1 + \xi_i \right) \\ & - \sum_{i=1}^n \beta_i \left(-y_i (w^T \phi(x_i) + b) + 1 + \frac{1}{\tau} \xi_i \right). \end{aligned} \quad (19)$$

通过引入核技巧, 式(18)的对偶问题可以表示为:

$$\begin{aligned} \min_{\alpha, \beta} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \beta_i) (\alpha_j - \beta_j) \phi(x_i)^T \phi(x_j) y_i y_j - \sum_{i=1}^n (\alpha_i - \beta_i) \\ \text{s.t.} \quad & \sum_{i=1}^n (\alpha_i - \beta_i) y_i = 0, \\ & \alpha_i + \frac{1}{\tau} \beta_i = s_i C, \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (20)$$

其中 $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ 是核函数, 常见的核函数为斯核函数: $k(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right), \sigma > 0$ 。

在非线性情形下求解出 Pin-FSVM 的参数 α, β 后, 决策函数为:

$$f(x) = \text{sign} \left(\sum_{i=1}^n (\alpha_i - \beta_i) y_i \phi(x)^T \phi(x_i) + b \right).$$

在本文中, 使用 CXVR 优化包求解提出的 Pin-FSVM 模型的参数, 求解过程如表 1 所示。

Table 1. The process of solving model parameters**表 1.** 求解模型参数过程

算法: 使用 CVXR 优化包求解 Pin-FSVM 的算法

初始化

1: 对指定的核函数和核函数的参数, 计算数据的内核距离;

2: 对于线性核函数, 直接计算 $k(x_i, x_j) = x_i^T x_j$;

3: 对于非线性情况, 使用高斯核函数, 可计算内积 $k(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$;

训练过程

4: 求解对偶问题线性情况式(13)矩阵形式的和非线性情况式(20)的解;

5: 将矩阵和数据传递给 CVXR 优化求解器求解出参数, 并利用式(15)求解出偏置项;

构建分类器

6: 对于给定的测试集, 使用求解出的支持向量和参数以及核函数对数据进行分类;

7: 计算测试集数据和支持向量的内核, 即可得决策函数为。

4. 基于 Pin-FSVM 的认购定期存款预测模型

4.1. 数据来源与指标设计

本文的数据来自 UCI 机器学习库(<https://archive.ics.uci.edu/>)公开的数据, 该数据集是一家葡萄牙银行机构 2008 年 5 月到 2010 年 11 月的直接电话营销活动, 收集的数据与葡萄牙一家银行机构定期存款的直接营销活动有关。通常, 银行需要与同一客户进行多次联系, 以便访问客户是否订阅产品(银行定期存款)[9]。主要目的在于促进现有客户的定期存款, 每条数据样本均有 20 个属性。本文的响应变量为客户是否会认购定期存款, 一共有两种可能的状态: 认购和不认购。综合考虑数据结构和指标特征, 在众多影响因素中主要考虑了以下 9 个解释变量对响应变量进行建模分析, 这些变量均与客户是否认购定期存款有高度相关关系[10]。变量具体说明如表 2。

Table 2. Data indicator variables**表 2.** 数据指标变量

变量类型	变量名	详细说明	取值范围
因变量	是否认购	客户认购的状态	0: 不认购 1: 认购
	年龄(age)	客户年龄, 反应客户所处年龄阶段。	20~80
解释变量	婚姻(marital)	婚姻状况: “离婚”, “已婚”, “单身”	0: 单身 1: 离婚 2: 已婚
	住房(housing)	分类变量: “未知”, “无”, “有”	-1: 未知 0: 无住房贷款 1: 有住房贷款
	贷款(loan)	分类变量: “未知”, “无”, “有”	-1: 未知 0: 无个人贷款 1: 有个人贷款

续表

活动(campaign)	活动期间为该客户进行的接触次数	取值为正整数
变化率(emp.var.rate)	就业变化率——季度指标	数值范围为: -3.4~1.4
价格指数(cons.price.idx)	消费者价格指数——月度指标	取值是数值, 范围为: 92.201~94.767
信心指数(cons.conf.idx)	消费者信心指数——月度指标	取值是数值, 范围为: -50.8~-26.9
月利率(euribor3m)	银行 3 个月利率	范围为: 0.635~5.045

(1) 年龄(age): 该指标是银行客户的年龄, 它反映的是客户存款能力。由于个人在不同年龄阶段创造财富的能力存在差异, 因此他们在消费、储蓄中所分配的份额就会有所不同, 银行针对个人的投资偏好、风险承受能力等因素, 对处于不同生命周期中的顾客分配了不同的商品和服务。不同的年龄段对是否认购定期存款有重要的作用, 客户的年龄反映客户的不同阶段, 对绝大多数人来说在 25~60 岁的人生创造价值的时期, 认购定期存款的概率要高一些。

(2) 婚姻(marital): 该指标是客户的婚姻状况, 用于反映客户的家庭情况之一。客户是否结婚与是否拥有稳定的经济来源有重要的关系, 结婚的客户由于婚后需要较大的经济开销, 单身的客户与已婚的客户相比, 经济压力要小一些, 所以更倾向于办理定期存款业务, 当然, 已婚的客户在后期经济足够稳定后也有很大的概率为孩子和以后的生活有保障认购定期存款。

(3) 住房(housing): 该指标是反映客户是否有住房贷款, 住房贷款是银行及其他金融机构向房屋购买者提供的任何形式的购房贷款支持, 通常以所购房屋作为抵押。住房贷款反映客户的经济情况, 有住房贷款的客户可能认购定期存款的概率要比没有住房贷款的客户低。住房贷款也从侧面反映了客户的家庭情况, 是一个比较重要的指标。

(4) 贷款(loan): 该指标是反映客户是否有个人贷款, 有个人贷款的客户一般情况下不会认购定期存款, 而没有个人贷款的客户有很大的可能认购定期存款, 这个指标可以比较直观的反映客户认购定期存款的情况。

(5) 活动(campaign): 该指标反映银行在这个活动期间与这个客户进行的接触次数(包括最后一次接触), 接触次数代表该客户是否认购定期存款的概率, 接触的越多代表客户越不可能认购, 该指标是正整数, 直观地反映了客户认购存款的需求。

(6) 变化率(emp.var.rate): 该指标反映的是就业变化率, 是季度指标。就业率是反映社会劳动力就业程度的指标。指在业人员占在业人员与待业人员之和的百分比。它反映全部可能参与社会劳动的劳动力中, 而实际被使用的人员比重。一段时期在业人员越多或待业人员越少, 则就业率就越高, 反之越低。计算和研究就业率, 实质上是要提高就业的经济效益。就业变化率是一个宏观指标, 是体现该时期整体社会的经济情况指标之一, 所以对客户是否认购定期存款有重要的作用, 就业变化率高, 定期存款就少一些, 反之, 就业变化率稳定, 定期存款就多。

(7) 价格指数(cons.price.idx): 该指标是消费者价格指数, 是月度指标。价格指数反映居民家庭一般所购买的消费商品和服务价格水平变动情况的宏观经济指标。CPI 及其分类指数还是计算国内生产总值以及资产、负债、消费、收入等实际价值的重要参考依据。价格指数与客户是否认购定期存款也有直接关系, 价格指数水平的上升, 意味着实际利率的下降, 如果存款利率保持不变, 价格指数持续上涨后, 甚至可能出现负利率, 使存款人遭受损失。所以为了继续吸引存款, 就要弥补价格指数上涨给存款人带来的损失, 银行就必须考虑合理的策略提高利率水平。所以, 利率水平与 CPI 具有同向变化的趋势, 价

格指数水平的变动成为影响利率水平的因素之一，其对客户是否定期存款也造成了影响。

(8) 信心指数(cons.conf.idx): 该指标是消费者信心指数，月度指标。消费者信心指数是一个衡量消费者对经济状况的信心程度的指标。消费者信心指数通常通过对消费者信心水平的调查来确定。调查通常涉及消费者对经济、就业、通货膨胀、个人财务状况等方面的看法。消费者信心指数越高，消费者越有可能在购买决策中表现出更积极的态度，反之亦然。它不仅可以反映出消费者对未来经济发展和个人财务状况的看法，也可以影响企业和政府的决策。消费者信心指数直接反映了整体是否认购定期存款的可能性，从宏观上体现了定期存款。

(9) 月利率(euribor3m): 该指标是欧洲银行 3 个月利率的每日指标，反映银行存款的月利率。月利率指以每个月为计算利息周期计算的利息(其中一个月内的利息额与存款、贷款本金的比率)利率按本金的千分之几表示。银行月利率的高低与客户是否认购定期存款成正比，月利率高，客户认购的定期存款就越多。

4.2. 定期存款预测模型的建立

在本节中进行实证研究，将提出的 Pin-SVM 应用于定期存款预测模型，实验都在 R4.3.1 中进行，在基于系统配置 Intel(R) Core(TM) i7-10700 CPU(2.90GHz)和 16 GB RAM 的 PC 上运行，使用优化工具包 CVXR 求解模型。关于模型参数的设置，对于 SVM、Pin-SVM、FSVM 和 Pin-FSVM，设置 $C = \{2^{-8}, 2^{-7}, \dots, 2^7, 2^8\}$ ；设置 Pin-SVM 和 Pin-FSVM 中 τ 的取值范围为 $\tau = \{0.1, 0.2, 0.5, 1\}$ 。非线性情况，考虑高斯核函数，评价指标为准确率(Acc.)。对于所有模型参数的选择，本文使用五折交叉验证法来搜索模型的最优参数。为了消除实验随机性的影响，进行 10 次数值实验，计算评价指标的平均值(Acc.)和标准差(sd)。考虑实际数据为非线性可分，所以建立高斯核函数下非线性可分模型，模型结果如表 3 所示。提出的模型提高了预测准确率，将其应用于客户是否认购定期存款预测模型可以带来更好的效果。

Table 3. Data set prediction results
表 3. 数据集预测结果

	SVM Acc. $\pm$ sd	FSVM Acc. $\pm$ sd	Pin-SVM Acc. $\pm$ sd	Pin-FSVM Acc. $\pm$ sd
数据集	0.6995 $\pm$ 0.025	0.672 $\pm$ 0.026	0.701 $\pm$ 0.027	0.713 $\pm$ 0.027

5. 结论

在当今的金融服务行业中，银行需要一种高效且准确的方法来识别其客户中哪些人可能倾向于认购定期存款。为了满足这一需求，本文提出了一种新的 Pin-FSVM 预测模型，模型结合了弹球损失函数和模糊隶属度的概念。通过将 FSVM 中的铰链损失函数替换为弹球损失函数，并结合模糊隶属度的概念，Pin-FSVM 不仅能够在存在噪声的情况下保持较高的预测准确率，还能更好地处理数据中的不确定性。这种结合使得 Pin-FSVM 在客户认购存款的预测任务中展现出了更好的性能。当 Pin-FSVM 被应用于银行客户认购存款的预测时，它能够准确地识别出那些更有可能选择认购定期存款的客户，增强了其预测准确率，为银行更好地识别和服务其客户群体提供了有力支持。这不仅有助于银行更精确地了解其客户群体的需求，还能帮助银行制定更有针对性的营销策略，从而提高客户满意度和忠诚度。

参考文献

[1] 秦玉芳. 定期存款利率下调 银行成本压力居高不下[N]. 中国经营报, 2022-05-02(B01).
[2] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297.

-
- <https://doi.org/10.1007/bf00994018>
- [3] Wang, H. and Zhang, Q. (2021) Twin k-Class Support Vector Classification with Pinball Loss. *Applied Soft Computing*, **113**, Article ID: 107929. <https://doi.org/10.1016/j.asoc.2021.107929>
 - [4] Durán Acevedo, C.M., Carrillo Gómez, J.K. and Albarracín Rojas, C.A. (2021) Academic Stress Detection on University Students during COVID-19 Outbreak by Using an Electronic Nose and the Galvanic Skin Response. *Biomedical Signal Processing and Control*, **68**, Article ID: 102756. <https://doi.org/10.1016/j.bspc.2021.102756>
 - [5] Qi, K. and Yang, H. (2022) Elastic Net Nonparallel Hyperplane Support Vector Machine and Its Geometrical Rationality. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 7199-7209. <https://doi.org/10.1109/tnnls.2021.3084404>
 - [6] Huang, X.L., Shi, L. and Suykens, J.A.K. (2014) Support Vector Machine Classifier with Pinball Loss. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**, 984-997. <https://doi.org/10.1109/tpami.2013.178>
 - [7] Lin, C.-F. and Wang, S. (2002) Fuzzy Support Vector Machines. *IEEE Transactions on Neural Networks*, **13**, 464-471. <https://doi.org/10.1109/72.991432>
 - [8] Jiang, X., Yi, Z. and Lv, J.C. (2006) Fuzzy SVM with a New Fuzzy Membership Function. *Neural Computing and Applications*, **15**, 268-276. <https://doi.org/10.1007/s00521-006-0028-z>
 - [9] 代雯月. 逻辑回归和支持向量机在客户信用分类中的应用[J]. 价值工程, 2023, 42(5): 139-141.
 - [10] 张利利, 郭淑妹, 马艳琴, 等. 基于数据挖掘技术的银行客户定期存款认购模型研究[J]. 数学的实践与认识, 2019, 49(21): 95-102.