Kernel Methods in Credit Risk Prediction - Using Support Vector Machines for Credit Analysis

Rebecka Algervik[*]

August 2024

Abstract

The credit risk of housing loans is undergoing dynamic changes in the global market- ing. With economic and market uncertainties, lenders are carefully assessing credit risk associated with housing loans. Factors such as borrower creditworthiness, em- ployment stability, and property market trends play pivotal roles in determining the level of credit risk. As financial institutions navigate these challenges, they employ advanced analytics and risk management strategies to reduce potential defaults and ensure the stability of their housing loan portfolios. This study analyzed the HMEQ dataset, comprising 3,364 observations and 12 in- put variables, to evaluate the performance of four types of Support Vector Machine (SVM) models: linear, polynomial, radial basis function (RBF), and sigmoid kernels. The data preprocessing steps included handling missing values, converting categor- ical variables into dummy variables, performing feature selection, addressing class imbalance by applying class weight balancing, and feature scaling. The dataset was split into training (70%) and testing (30%) sets. Model performance was assessed us- ing metrics such as accuracy, precision, recall, false positive rate, F1 score, specificity, AUC values, and cross-validation. The results indi- cated that the Polynomial kernel achieved the highest accuracy and cross-validation scores, demonstrating its effectiveness in this context. However, considering the AUC values, the RBF model emerges as the most suitable model for this dataset. Additionally, balancing class weights effectively addressed the issue of data imbalance.

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: rebecka.algervik@gmail.com. Supervisor: Taras Bodnar and Jan-Olov Persson.