Effect of the Regularization Parameter (C) on Margin Width and Generalization in SVM

Introduction

Support Vector Machines (SVMs) are widely used machine learning algorithms known for their strong theoretical foundation and excellent generalization performance. A key factor influencing the behaviour of SVMs is the regularization parameter, denoted as **C**, which controls the trade-off between maximizing the margin and minimizing classification errors. The choice of C directly affects the width of the margin and the model's sensitivity to training data. A small value of C allows more misclassifications, resulting in a wider margin and improved robustness to noise. In contrast, a large C penalizes errors more heavily, producing a narrower margin that closely fits the training data. This trade-off reflects the fundamental bias–variance dilemma in machine learning. Understanding how C influences margin width and generalization is essential for building reliable and well-performing SVM models. This study investigates the impact of different values of C on model behaviour and predictive performance.

What is Support Vector Machine?

Support Vector Machine (SVM) is a robust and widely used supervised machine learning algorithm designed for both classification and regression problems. Its core objective is to identify an optimal decision boundary, known as a hyperplane, that separates data points of different classes with the maximum possible margin. The margin represents the distance between the hyperplane and the closest data points from each class, which are referred to as support vectors. By maximizing this margin, SVM enhances its ability to generalize well to unseen data and reduces sensitivity to noise. SVM is effective for both linearly separable and non-linearly separable datasets through the use of kernel functions. Popular kernels such as linear, polynomial, radial basis function (RBF) and sigmoid enable SVM to implicitly map data into higher-dimensional feature spaces where class separation becomes feasible. This kernel trick allows SVM to solve complex, non-linear problems without explicitly increasing computational complexity. SVM performs particularly well in high-dimensional spaces and in scenarios where the number of features is large relative to the number of samples. Additionally, regularization plays a crucial role in controlling model complexity and preventing overfitting. Due to its strong theoretical foundations and flexibility, SVM has been successfully applied in areas such as text classification, bioinformatics, image processing, financial analysis and medical diagnosis.

The Role of the Regularization Parameter (C)

The regularization parameter C plays a critical role in controlling the behaviour and performance of Support Vector Machines by explicitly regulating the trade-off between margin maximization and empirical risk minimization. From an optimization perspective, C acts as a penalty coefficient on the slack variables, determining how strongly misclassified or margin-violating samples are penalized in the objective function. A smaller value of C imposes stronger regularization, allowing the model to tolerate a greater number of classification errors in order to maintain a wider margin. This encourages smoother decision boundaries, increases bias and reduces variance, which can be particularly beneficial in noisy or high-dimensional datasets. Conversely, a larger value of C weakens regularization by assigning a high cost to misclassifications, compelling the SVM to fit the training data more closely. This results in a narrower margin and a more complex decision boundary that is sensitive to individual data points and outliers. Although this may improve training accuracy, it significantly increases the risk of overfitting and reduced robustness on unseen data. Consequently, the choice of C directly governs the bias–variance trade-off, margin geometry and generalization ability, making it one of the most influential hyperparameters in SVM model design.

Dataset

The dataset used in this study is the Breast Cancer Wisconsin (Diagnostic) dataset, a well-known benchmark dataset commonly employed for evaluating classification algorithms. It consists of 569 patient samples, each labelled as either malignant or benign, making it a binary classification problem suitable for studying model generalization. The dataset contains 30 real-valued numerical features computed from digitized images of fine needle aspirate (FNA) biopsies of breast masses. These features describe various characteristics of cell nuclei, including radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry and fractal dimension. For each characteristic, statistical measures such as mean, standard error and worst (largest) values are provided, offering a comprehensive representation of the data. The dataset is clean, with no missing values, which eliminates the need for extensive preprocessing and allows the analysis to focus on model behaviour rather than data quality issues. In this study, the dataset is divided into training and testing subsets to evaluate how different values of the SVM regularization parameter affect margin width and generalization performance.

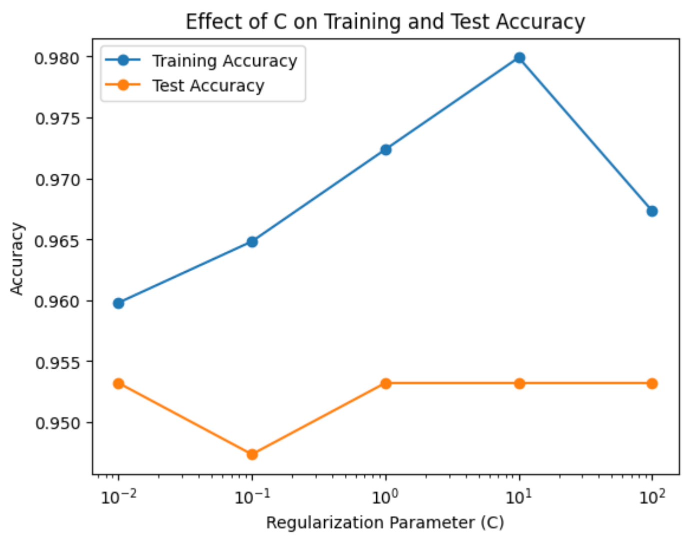

Figure 1: Effect of the Regularization Parameter (C) on Training and Test Accuracy.

Experiment Analysis and Discussion

The experimental results provide clear empirical evidence of how the regularization parameter C influences margin width, classification accuracy and generalization performance in Support Vector Machines. By systematically varying C and analysing the resulting plots and confusion matrix, the underlying bias–variance trade-off governed by regularization can be clearly observed.

The margin width plot demonstrates a strong inverse relationship between C and the width of the separating margin. At very low values of C (e.g., C = 0.01), the margin is exceptionally wide, indicating strong regularization. In this regime, the SVM allows several margin violations in order to maintain a simple and smooth decision boundary. Such behaviour reflects a high-bias, low-variance model that prioritizes generalization over training accuracy. As C increases, the margin width decreases rapidly, showing that the model increasingly penalizes misclassified samples and margin violations. At large C values (C = 10 and C = 100), the margin becomes extremely narrow, suggesting that the classifier is tightly fitting the training data and becoming more sensitive to individual observations and potential noise.

- Small C → large margin
- Large C → narrow margin

This confirms that stronger regularization encourages wider margins, which generally leads to better generalization.

This behaviour is further reinforced by the **training and test accuracy plot**. Training accuracy consistently increases with higher values of C, reaching its peak at larger C values where misclassifications are heavily penalized. This confirms that weaker regularization enables the model to better fit the training data. However, test accuracy does not follow the same increasing trend. Instead, it remains relatively stable across moderate values of C and shows no meaningful improvement at higher C levels. The growing gap between training and test accuracy at large C values is a clear indication of overfitting, where improved training performance does not translate into better generalization on unseen data. Conversely, very small values of C slightly reduce both training and test accuracy, reflecting underfitting due to excessive regularization. These results highlight that optimal generalization is achieved at intermediate C values, where the bias–variance trade-off is balanced.

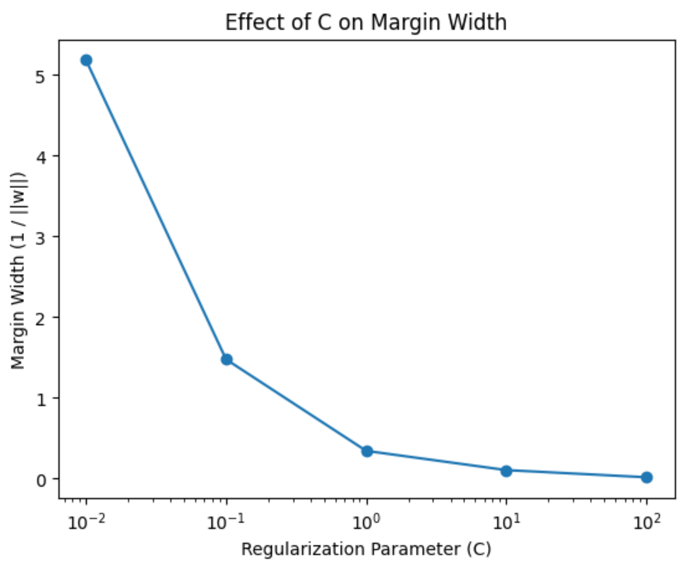

Figure 2: Effect of the Regularization Parameter (C) on Margin Width.

The confusion matrix for C = 1 provides a more granular evaluation of classification performance beyond accuracy alone. The model correctly classifies the majority of both classes, with 57 true negatives and 106 true positives, indicating strong predictive capability. Only 7 false positives and 1 false negative are observed, resulting in a very low misclassification rate. The minimal number of false negatives is particularly significant in the context of medical diagnosis, as it suggests that the model rarely misidentifies malignant cases as benign. This reflects the robustness of the decision boundary produced at C = 1, where regularization is sufficient to prevent overfitting while still capturing the underlying class structure. The confusion matrix therefore confirms that the selected C value not only yields high overall accuracy but also achieves a desirable balance between.

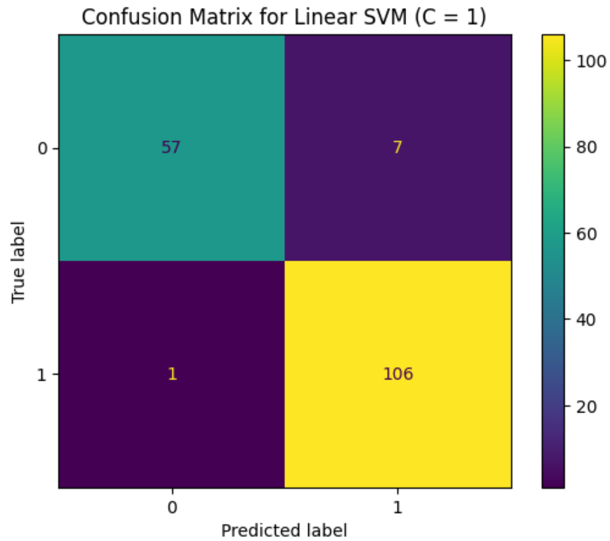

Figure 3: Confusion Matrix for Linear SVM with Moderate Regularization (C = 1).

Taken together, the plots and confusion matrix collectively demonstrate that the regularization parameter C plays a decisive role in shaping SVM behaviour. Small C values lead to wide margins and strong generalization at the expense of accuracy, while large C values result in narrow margins and overfitting. The experimental findings validate theoretical expectations and show that moderate regularization (C ≈ 1) offers the best compromise between margin width, predictive accuracy and generalization performance. This emphasizes the importance of careful hyperparameter tuning when applying SVMs to real-world classification problems.

Conclusion

This study examined the effect of the regularization parameter C on margin width and generalization performance in Support Vector Machines. The experimental results showed that smaller values of C produce wider margins and simpler decision boundaries, which enhance robustness but may lead to underfitting. In contrast, larger C values yield narrower margins and higher training accuracy at the cost of increased sensitivity to noise and overfitting. The analysis of training and test accuracy confirmed the presence of a clear bias–variance trade-off governed by C. The confusion matrix further demonstrated that moderate regularization achieves a strong balance between sensitivity and specificity. Overall, the findings highlight that selecting an appropriate value of C is crucial for building reliable and well-generalized SVM models.

References:

1. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.
2. Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.
3. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222.
4. Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.".

External Links:

1. Github : arsal1122/Effect-of-the-Regularization-Parameter-C-on-Margin-Width-and-Generalization-in-SVM