Support Vector Machine (SVM) Classification on Fashion-MNIST

# 1. Introduction

Support Vector Machines (SVM) are supervised learning algorithms primarily used for classification and regression tasks.

The central idea is to find an optimal hyperplane that separates data points of different classes with the maximum margin.

For datasets that are not linearly separable, SVM uses kernel functions to map the data into a higher-dimensional space where a separating hyperplane can be found.

# 2. Dataset Used

The dataset used in this experiment is the **Fashion-MNIST** dataset. It consists of 60,000 training images and 10,000 test images. Each image is a 28x28 grayscale representation of clothing items across 10 categories such as T-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot. This dataset is widely used as a more challenging replacement for the original MNIST digits dataset.

# 3. How SVM Works

SVM works by finding the decision boundary that best separates different classes. The key concepts include:

• Hyperplane: A decision boundary that separates classes.  
• Margin: The distance between the hyperplane and the closest data points from each class. SVM maximizes this margin to improve generalization.  
• Support Vectors: The data points that lie closest to the decision boundary and directly influence its position.

# 4. Kernels in SVM

SVM can handle non-linear data through kernel functions, which implicitly map input data into higher dimensions:  
• Linear Kernel – Best for linearly separable data.  
• Polynomial Kernel – Captures polynomial relationships of degree d.  
• Radial Basis Function (RBF) Kernel – Popular choice, effective for non-linear data.  
• Sigmoid Kernel – Similar to neural networks, though less commonly used.  
Other kernels can also be defined, depending on the data characteristics.

# 5. Hyperparameters in SVM

The key hyperparameters tuned in SVM are:  
• C: Regularization parameter. A small C allows a wider margin with more misclassifications, while a large C aims for perfect classification but may overfit.  
• Gamma: Defines the influence of a single training example in RBF/Polynomial kernels. Low gamma means far-reaching influence, high gamma means close influence.  
• Kernel: The kernel function type (linear, rbf, poly, sigmoid).

# 6. Metrics Used

The following metrics were used to evaluate the performance of the SVM models:  
• Accuracy: The proportion of correctly classified samples.  
• Confusion Matrix: Provides detailed insights into classification performance across different categories.  
• Accuracy vs. Hyperparameters Plot: Shows how accuracy varies with different kernel functions and values of C.

# 7. Results

Experiments were conducted using different kernels (Linear, RBF, Polynomial, Sigmoid) and hyperparameter values for C (0.1, 1, 10). Key findings include:  
• The Linear kernel achieved good results but struggled with non-linear patterns.  
• The RBF kernel provided the best overall accuracy, as it can model complex boundaries.  
• Polynomial kernels captured non-linear patterns but required careful tuning of degree and parameters.  
• The Sigmoid kernel generally underperformed compared to others.  
Plots of accuracy vs. kernels and confusion matrices highlighted these performance differences.

# 8. Conclusion

Support Vector Machines are powerful models for classification tasks, especially when equipped with kernel functions. On the Fashion-MNIST dataset, the RBF kernel provided the highest accuracy, demonstrating its ability to capture complex class boundaries. Hyperparameters like C and Gamma significantly impact performance and require careful tuning. Overall, SVM proved to be an effective algorithm for image classification in this context.