Report: Evaluation of K-Nearest Neighbor Classifier for Image Classification

Introduction: The purpose of this report is to assess the performance of a K-Nearest Neighbor (KNN) classifier on the CIFAR-10 dataset, a widely used benchmark for image classification tasks. The KNN algorithm is chosen due to its simplicity and effectiveness, particularly for small to medium-sized datasets.

Methodology:

1. **Data Acquisition and Preprocessing:** The CIFAR-10 dataset, consisting of 60,000 32x32 color images across 10 classes, is obtained from an online source. To reduce computational complexity and memory usage, a subset of 1000 training and 100 test examples is selected.
2. **K-Nearest-Neighbor Class Implementation:** A custom K-Nearest-Neighbor class is developed to encapsulate the functionality of the KNN algorithm. This class includes methods for training the classifier, computing distances between data points, and predicting labels for test examples.
3. **Cross-Validation:** To determine the optimal value of k (number of nearest neighbors) and the best distance metric (Euclidean or Manhattan), 5-fold cross-validation is performed. For each combination of parameters, the average accuracy is calculated over the validation folds.
4. **Performance Visualization:** The plot-comparison function is utilized to visualize the performance of the KNN classifier with different distance metrics and k values. This aids in understanding the impact of these parameters on classification accuracy.
5. **Model Evaluation:** Finally, the trained KNN classifier is evaluated on the test set using the best k and distance metric. The test accuracy is computed to assess the generalization performance of the model.

Results and Discussion: The cross-validation results indicate that both Euclidean and Manhattan distance metrics are effective for image classification, with slight variations in performance. The

optimal k value varies depending on the distance metric used. The test accuracy achieved by the selected model demonstrates its effectiveness in accurately classifying unseen data.

Conclusion: In conclusion, the evaluation of the KNN classifier on the CIFAR-10 dataset highlights its suitability for image classification tasks. By considering various parameters such as k and distance metric, we can fine-tune the model to achieve optimal performance. The insights gained from this evaluation can guide the selection and configuration of machine learning models for similar tasks in the future.