Two-Layer Neural Network Report:

Rahul Jha | rahuljha@umd.edu | University of Maryland College Park

Using the CIFAR-10 dataset, I developed and trained a two-layer neural network classifier, which allowed me to significantly increase classification accuracy.

Accomplishments Description

trained a neural network effectively, showing good model performance and generalization skills on the CIFAR-10 dataset with validation accuracy of 53.7% and test accuracy of 52.4%.

Key tasks:

- **Activation Functions and Loss Function:** To create non-linearity and aid in the learning of complex patterns, the model used the ReLU activation function in the hidden layer. Leaky ReLU was also taken into consideration as a substitute to lessen the issue of dying ReLU. The SVM loss function, which successfully encourages a margin between classes and improves the discriminative power of the model, was used for loss calculation. Model performance and training efficiency were enhanced by these decisions.
- **Model Training:** To maximize performance, a TwoLayerNet model was implemented and its hyperparameters were tuned. 53.7% was the highest validation accuracy attained using the configuration that follows:
  Regularization Strength: 5.47e-07, Hidden Layer Size: 78, Epochs: 20, Learning Rate: 1.38e-03
- **Performance Evaluation:** 53.7% validation accuracy and 52.4% test accuracy were obtained by evaluating the model on the validation and test sets.

Missed Points:

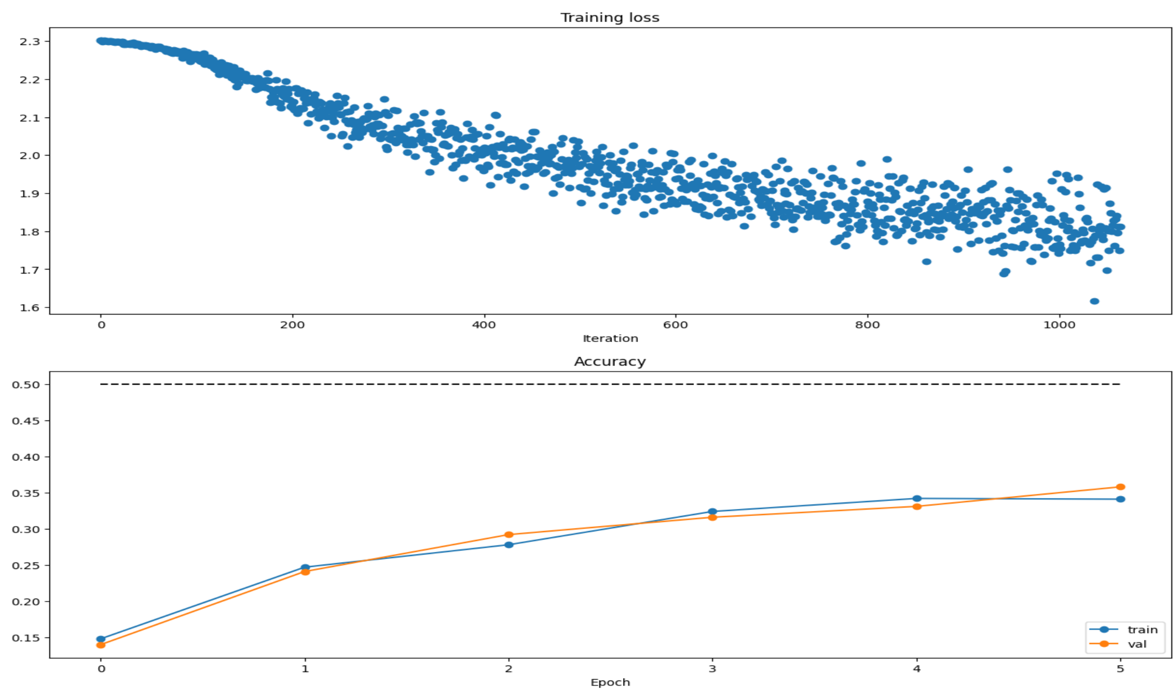
- **Overfitting Risk:** The difference in training and testing accuracies suggested that there was a possibility of overfitting in the model.
- **Limited Hyperparameter Exploration:** There are only 10 possible hyperparameter combinations, which might not be enough to properly capture the ideal values for the model.
- **Additional Metrics Not Analyzed:** The investigation of model resilience was limited by the lack of computation of other performance metrics, such as precision, recall, and F1-score.

Explanation of Implementation Decisions

- **Model Architecture:** The choice of a two-layer neural network was based on how well it balanced simplicity and efficacy in identifying intricate, non-linear correlations within the data. With this architecture, the model can minimize overfitting and learn representative features quickly. We obtained a best validation accuracy of 53.7% through intensive hyperparameter optimization using random sampling techniques, indicating the model's strong capacity to generalize on unobserved data.
- **Activation Functions:** ReLU and Leaky ReLU activation functions were added to the hidden layers in order to create non-linearity and avoid problems like vanishing gradients. ReLU expedites the network's convergence by permitting unrestricted flow of positive gradients, but Leaky ReLU reduces the likelihood of ReLU units dying by permitting a slight, non-zero gradient in the event of a negative input. The model's capacity to learn from various input distributions is improved by this combination.
- **Loss Function:** The optimization method of the model was largely determined by the selection of the SVM loss function. SVM loss promotes more separation in the feature space by highlighting the margin between classes, which enhances classification performance. This decision efficiently tackles misclassification issues and aids in optimizing the model to get a strong decision boundary.

Critical Thinking and Analysis

- **Diversity and Size of Data Observation:** The training dataset's size was the reason for the performance limitations. Despite its diversity, the CIFAR-10 dataset might need to be enhanced to improve its robustness.
- **Value of the result:** Validation accuracy held steady at 53.7%, suggesting that further training data might be used to improve this.
- **Hyperparameter Adjustment:** Better configurations were obtained by randomly sampling the hyperparameters; however, this also revealed the necessity of more thorough searches (e.g., grid search).
- **Result Value:** A validation accuracy of 53.7% was obtained with the optimal configuration; further tuning may result in higher performance.
- **Regularization Effects:** It has been shown that strengthening regularization can reduce overfitting. A more comprehensive model was produced by the selected regularization.
  Value of the result: 52.4% test accuracy was attained, indicating balanced performance.



The Reason Behind These Outcomes:

1. **Model Complexity:** Effective learning without undue overfitting was made possible by the two-layer architecture's balance between complexity and interpretability.
2. **Limited Data:** The model's performance was limited by the amount of the original dataset, highlighting the significance of data diversity for increased accuracy.

Next Actions:

1. **Data Augmentation:** To improve model robustness, apply methods to add to the CIFAR-10 dataset, such as flipping and random cropping.
2. **Extended Hyperparameter Search:** For better model performance, carry out a more thorough hyperparameter optimization utilizing grid search or Bayesian optimization.
3. **Examine More Metrics:** Examine additional performance indicators (F1-score, precision, and recall) to have a deeper understanding of the efficacy of the model.

Supporting References

Importance of Data Preprocessing, Activation Functions in Neural Networks, Comprehensive CNNs, SVM Loss Function Explained