Vision Transformer on CIFAR-10

We substituted the standard linear layers with NdLinear modules both during the patch projection step and the MLP head in our custom Vision Transformer (ViT) implementation. This modification preserves the spatial data of image patches through dimension-specific transformations rather than data flattening, resulting in a more parameter-efficient model. By contrasting this NdLinear-boosted ViT with a baseline model of standard nn.Linear layers, we observed dramatic parameter reductions while maintaining—at points even surpassing—classification performance on the CIFAR-10 dataset.

The impact of including NdLinear was noticed in more balanced training dynamics, with a steadily decreasing loss and rising validation accuracy with each epoch. This new approach aligns well with the project's requirement for a point of comparison, and it distinctly indicates that preserving multidimensional forms can not only ensure maximum computational efficacy but also enhance the representational power of the model. Overall, the NdLinear-inspired ViT is strong evidence of the potential of structured linear transformations to drive performance and efficiency in image classification tasks.

MLP on Tabular Data

On the tabular data experiment, we used an MLP on the UCI Breast Cancer dataset in place of regular dense layers using NdLinear layers. Our MLP model, with the use of the multi-dimensional transformation ability of NdLinear, is able to capture and preserve the structural relationships present in the input features more effectively. This design choice resulted in a model that required significantly fewer parameters than an ordinary MLP realized using nn.Linear layers, but with comparable—and in some cases, superior—classification accuracy.

The experiment results showed that the NdLinear-enhanced MLP converged reliably with a smooth reduction in training loss, and the corresponding validation accuracy indicated robust performance. The alignment with the project outline is highlighted through the benchmark comparison: our results confirm definitively that the parameter efficiency facilitated by NdLinear does not come at the expense of performance. Instead, the structured nature of NdLinear is to support steady predictive outcomes, and thus it is an incredibly effective and creative alternative to optimizing MLP models on tabular data.

BERT for Text Classification

In our BERT-based text classification model, we added NdLinear layers to the classifier head to further transform the structured embeddings from the BERT backbone. Replacing the regular feed-forward dense layers with NdLinear allowed the model to maintain the token-level relationships inherent in the input while significantly reducing the parameter footprint. Benchmark comparisons against a regular classifier head revealed that the NdLinear-extended BERT exhibited better convergence behavior and better classification metrics on the SST-2 benchmark.

The results demonstrated that the model based on NdLinear made well-balanced predictions as indicated by a well-behaved confusion matrix and stable training curves. By preserving the multi-dimensionality in the output layers, NdLinear facilitated the ability of the model to distinguish sentiment classes better without overparameterization. This creative merge emphasizes the functional benefits of NdLinear in enhancing transformer models, ideally as per the project goal of emphasizing performance improvements and parameter effectiveness through comparison benchmarks on an open-source data set.