Image Classification with MLP and CNN

Name : Yachi Darji
Data Science, Illinois Institute of Technology; Chicago, IL, United States

Abstract : This paper shows comparison between Multi-layer perceptron (MLP) and Convolutional Neural Network for image recognition task on CIFAR-10 data. Models are implemented using PyTorch. As a result, CNN works better than MLP. Overall proved, CNN is good choice for image data.

I. Introduction

Image recognition is an application of machine learning (ML) (or machine learning technology) that enables software and devices to identify objects, places, people, writing and actions in digital images or videos.[?] Image recognition is advancing rapidly, paving the way for more sophisticated applications across numerous industries and use cases such as autonomous vehicle, facial recognition, social media management, and moderation.[?] This task is tough, as classes represent variations(pose, lighting, background, noise) within class and similarity among different classes. Here I address this problem using pure PyTorch and CIFAR - 10 Data which contain 60,000 RGB images (32×32 pixels) with ten distinct classes. Furthermore, implemented two architectural models 1.Multi-Layer Perceptron (MLP) 2. Convolutional Neural Network (CNN) My goal is to compare the performance of both models under different training conditions, and I ask 1. Does CNN model outperform MLP? 2. How do different learning rates impact accuracy and runtime. My repository setup for this assignment :

Artifacts\ : Plots + CSV log
src\data.py : download\load data
src\train.py : Training + Validation
src\utils.py : Helper Function
src\exp.py : Batch size + learning rate + momentum
src\models.py : MLP + CNN
main.py : run models + experiments
requirement.py : software list of dependencies

II. Methodology

CIFAR-10 splitted into 50k training and 10k used as validation set. Here are two model architectures with different inductive biases 1. MLP : is a name for a modern feedforward neural network consisting of fully connected neurons with nonlinear activation functions, organized in layers, notable for being able to distinguish data that is not linearly separable.[?] Each image (32×32) with 3 color channels result vector is $3 \times 32 \times 32$. The model flattened 2d images into 1d vectors followed by linear mapping 3072 units to 512 hidden units with ReLU activation. Furthermore, Linear mapping from 512 units to 256 units after ReLU activation. For the Output Layer, 256 Units to 10 class mapping. ReLU works as an activation function, to introduce non linearity and helps neural networks to learn difficult patterns instead of just combinations. The output logits are further passed to the loss function to compare it with true classes. It's fast and simple but underperforms on image data, because while flattening it destroys crucial information from 2d data (shapes, proximity).

CNN : is a type of feedforward neural network that learns features via filter (or kernel) optimization.[?] The key idea is instead of looking at the whole image just look at the patch of image and step by step combine those patches into the high level pattern. For CIFAR -10 my stack is the Input layer which reads image data ($32 \times 32 \times 3$). The convolutional layer applies numerous small filters (3×3) over input to produce feature maps. It helps to learn local features(edge, texture). The Activation Layer applies ReLU to approximate complex patterns. Pooling Layer helps to reduce volume which makes computation fast and prevents overfitting. After conv + pooling, result features are flattened to one dimensional vector. A fully connected layer takes input from the previous layer and computes the final classification. The resulting logits from a fully connected layer are fed into a loss function for training.

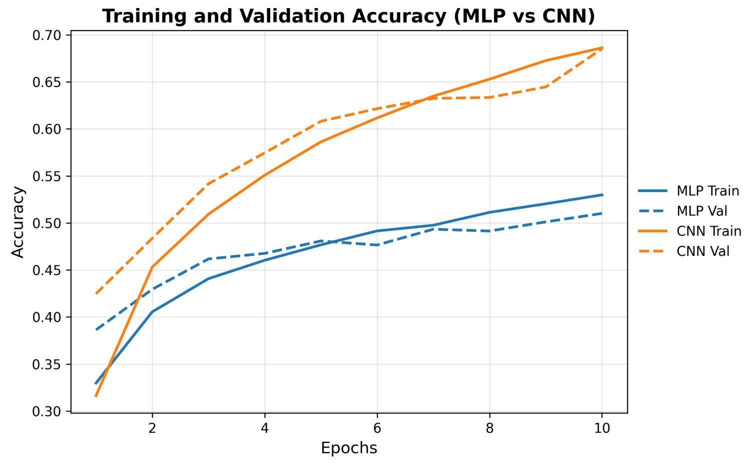

Fig. 1. CNN vs MLP training accuracy

Fig 1 shows comparison of CNN and MLP training and validation accuracy in ten epochs.

MLP Training accuracy starts about 0.32 increases slowly and flattens out around 0.52. In contrast CNN starts at 0.31 rises sharply and peaks to 0.68. This shows CNN captures far better spatial pattern than the MLP.

MLP validation accuracy increased gradually and settled around 0.51. While CNN validation accuracy stabilizes higher -0.68 aligns with training accuracy. This shows CNN can better generalize than MLP. Afterwards MLP struggles to improve accuracy.

CNN consistently outperforms MLP in training and validation accuracy which shows CNN converges faster. There is a clear gap in training and validation of MLP shows underfits because it fails to capture local patterns. In contrast, CNN training and validation stayed close, which shows less risk of overfitting.

I can conclude that CNN is superior for this problem while MLP has limited abilities. Even if training them longer will not improve MLP.

III. Experimental Design & Results

Fig 3 shows How different learning rate impacts accuracy.

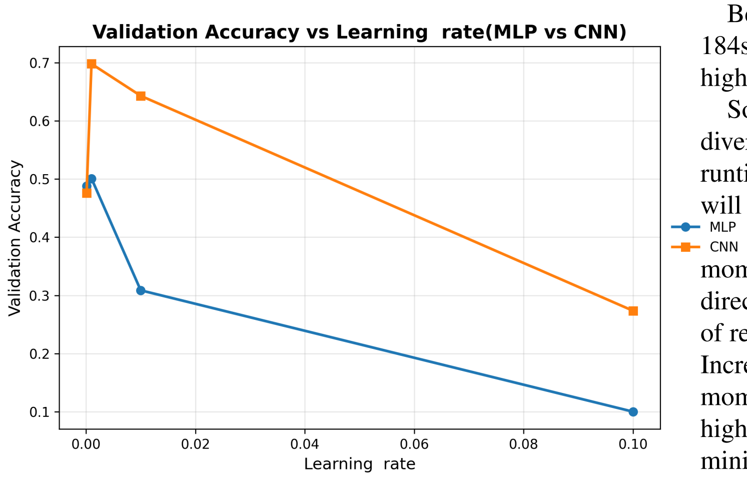

Fig. 2. Accuracy vs Learning Rate

The learning rate is a hyperparameter that controls the speed at which a neural network learns. It' s a measure of how much the weights of the neural network are updated each time the model is trained.? Every time model optimizes it updates parameter based on result. High learning rate updates weight fast, but low learning rate updates weights slowly. (e.g., climbing down hill with safe speed of descent).

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla L(\mathbf{w}_t)$$

$$\eta = \text{learning rate}, \quad \nabla L(\mathbf{w}_t) = \text{slope of the loss function}$$

Here, I have tested learning rate 0.0001 to 0.1 with validation accuracy. CNN stays high for all parameters that shows its already better model than MLP. For small LR 0.0001,both models perform average, but CNN outperforms MLP and reaches 0.7 whereas MLP reaches to 0.5.

That states, small LR doesn't let model learn in every steps. Because model is updating parameter slowly and it will take too long to reach convergence. But that does'nt prove high LR is the best as Validation accuracy starts to drop as LR increases. Model is looking for optimal learning rate. Model changes its internal parameter at every training step to improve performance.The size of this change highly depend on learning rate. Hence, Optimal learning rate is important for high performance.

TABLE I
Val Acc and Runtime across LR for CNN and MLP

LR	CNN Val Acc	CNN Runtime	MLP Val Acc	MLP Runtime
0.0001	0.4761	229.71	0.4883	102.21
0.001	0.6982	184.12	0.5010	98.09
0.01	0.6432	229.65	0.3087	93.74
0.1	0.2736	201.32	0.1000	98.05

CNN runs longer on all LRs than MLP as it has more convolutional operations to run. This extra runtime yields much higher accuracy. Runtime doesnt vary much across all LRs But accuracy collapses at higher LR. That proves, LR affects accuracy much more than runtime.

Best LR is 0.001, as CNN accuracy is 0.69 and runtime is 184sec as well as for MLP 0.50 and 98 sec. Both results are higher across all other rates.

So when the learning rate is too high 0.1, can cause model divergence leads to low validation accuracy 0.27 with regular runtime of 201 sec. This means model will not converge, it will not be able to generalize to new data.

Learning rate optimization is highly dependent on momentum. LR updates step size while momentum gives direction of update by giving past gradient information. Instead of relying only on current state it considers previous state info. Increasing the momentum can lead to faster convergence. Low momentum can stall training at minor local minima, while high momentum can accidentally skip over significant local minima.?

IV. Future Work

For future work, I would like to use regularization methods drop out or batch normalization for more generalize it. As well as, like to work on large and different datasets using pre-trained models for image classification tasks.

V. Conclusions

CNN shows high training and validation accuracy even better generalization with different parameter like batch size, learning rates and momentum. MLP's runtime depend on parameters like learning rate while CNN is stable. Overall, CNN is superior for image classification .