Problem 1 : LeNet, Conv Nets and Residual Blocks

Model Comparison Analysis

Model	Loss	Train	Test	Notes
Basic LeNet	1.2509	0.5477	0.5760	Baseline performance, consistently improved across all 25 epochs
LeNet Scratch	1.2417	0.5449	0.5777	My model from scratch performed better than the PyTorch one, however slight fluctuations in later epochs mean LR could be tuned slightly more
LeNet ReLU MaxPool	1.2894	0.5607	0.6060	Changing to ReLU and MaxPool improved overall performance slightly but less stable training
LeNet ReLU MaxPool 3Conv	0.9509	0.6914	0.7042	Adding a 3rd convolutional layer improved accuracy by 10% but there was less correlation between loss/train/test acc suggesting less consistency in training
BNLeNet	0.5243	0.8152	0.7919	Batch Normalization improved accuracy by another 10% and training was slightly more stable
BNLeNet ResBlock	0.3531	0.8747	0.8369	Batch Norm with a residual block was the best performing model on CIFAR, with stable training, lowest loss / accuracy even with a higher learning rate.

There were a few bugs in the starter code that I needed to manage and the D2L package refused to work so I had to use Claude to generate a minimal replacement D2L wrapper that allowed me to run the starter code.

Problem 2 : RNN and LSTM | Spanish-English Translation

Model Comparison Analysis

Model	Train Loss	Val Loss	BLEU Score	Notes
RNN	5.9213	6.1814	3.79	Validation Loss increased in later epochs, suggesting overfitting. Example output was gibberish with repeated words and no meaningful improvement in output.

LSTM	3.7842	4.6421	5.12	Training and validation loss both improved our all 5 epochs. Example output improved and in the final epoch there were some similar words and vaguely similar sentence patterns. However the output still wasn't good English.

RNN vs LSTM

LSTM performed better with loss and BLEU score likely due to its ability to store context weight of important words in longer term memory for longer sequences. However, both models were not particularly effective as the industry standard BLEU score is 40+ and Humans communicate at 70+.

Problem 3 : My RNN Implementation | Spanish-English Translation

Model	Train Loss	Val Loss	BLEU Score	Notes
My RNN	5.7678	6.2437	2.84	Validation Loss was actually highest in the first epoch and fluctuated for the remainder, suggesting the RNN overfit and would not generalize well to new data.

My RNN vs Torch RNN

My implementation performed better in terms of training loss but worse in terms of validation loss than the torch implementation. The BLEU score was marginally lower. The similar training curves suggest my implementation was correct but the RNN encoder-decoder is not an effective model for seq2seq translation.