**Introduction and Goal**

This report focuses on the cifar10 benchmark in Keras. We train a model on the first 3200 training images with 800 for validation and use first 2000 test images fir final validation. This dataset includes full color images ten different classes:

"airplane","auto","bird","cat","deer","dog","frog","horse","ship","truck"

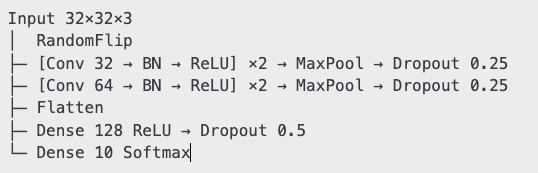
Our goal is to accurately classify these images using a convolutional neural network.

**Preprocessing**

Our preprocessing is done with a light touch. We first rescale the pixel values from 0-225 to 0-1, and then at train time we randomly flip the image to improve effectiveness and abstraction.

**Model Development**

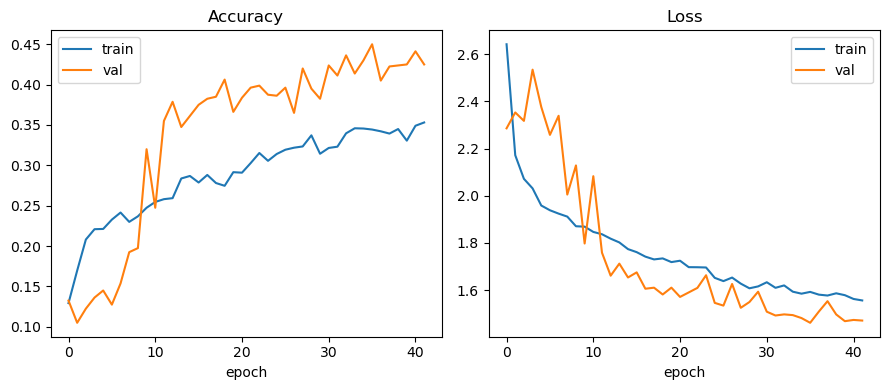
Our model Is built along the following design.

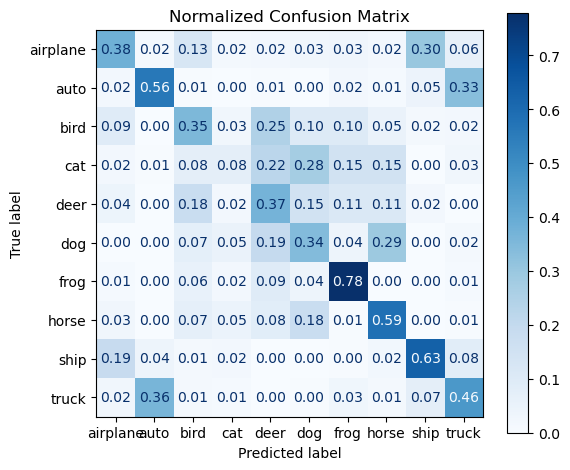


We use the Adam Optimizer and have early stopping monitoring validation loss with patience set to 6. Full model code is available in the attached Jupiter notebook.

**Performance**

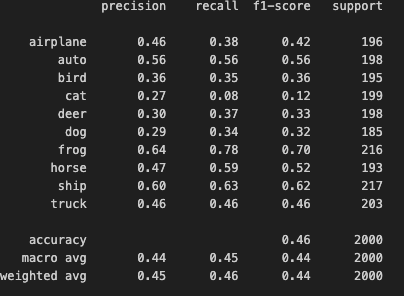
We see that the model begins to plateau after around 30 epochs, but continues slower growth thereafter. This indicates that in future iterations, a higher learning rate, or more epochs may add value. That said, I believe the result is satisfactory. In the end, we reach a test accuracy of 45%. This is inline with expected accuracy from our train-data as shown below.



When we go a level deeper to look at the confusion matrix, it is clear that our model is better at classifying some objects than others. It is very excellent with frogs, getting them right 78% of the time, and does well with horses, ships, and automobiles all around 55-60%. On the other hand, it is unable to discern cats with an accuracy of only 8%. As shown in the normalized confusion matrix below

Another interesting trend we see emerge is that some items are harder to distinguish between than others. For example, Airplanes get classified as ships 30% of the time, and autos get classified as trucks 33% of the time. This is significantly higher than the error rate of most other pairs hovering below 7%.

A more directly interpretable way to look at this model’s performance is to use precision recall and F1 score for each group as shown below



This reaffirms our knowledge that the model does well with autos, frogs and ships while performing much worse with cats.

**Conclusion**

Altogether, this model does an ok job predicting images. With more creativity an exploration, this performance could continue to be balanced and improved across classes.