Assignment 4

Yang Zhao



Among the results, models with embedding layer from scratch perform best. Embedding layer with masking do helps when the training size is small, while this improvement disappears when increase the training size. Increasing the training size can significantly increase performance. But the pretrained word embedding model does not perform as good as I thought. Possibly because the IMDB example already contains enough observations to train a model which is more closely fit the example. The whole accuracy is lower than what’s in the textbook, which may be due to the short cutoff (150 words) of review and lower max token length (10000).

Google colab web link: <https://github.com/zhaoyang1209/yangzhao_64061.git>