**Minutes**

17/03/23:

* Discussed randomness of training:
  + With weights being updated in the training loop, conducting an experiment multiple times can give different performances on validation/test data due to randomness of shuffling, order of sequences and gradient updates.
  + Quite often in large-scale assessments, maybe 100 models are trained and a gaussian distribution can be plotted to see how stable the model is. A model with better stability will have less spread.
  + I have saved ‘pth’ files containing my final trained weights, experiments can be repeated easily, but entire training and tuning of models could give slightly different results if this was repeated.
    - To do: put .pth files on github for ease of use.
* Discussed evaluation metrics:
  + Current confusion matrix and metrics as an average over sequences look good.
  + Looking at ROC curve: where the threshold value (currently 0.5) can change. For example, if we want to guarantee finding every disordered amino acid this threshold could be lowered (gives better recall).
    - Sklearn have a function that can take the raw probabilities and plot these curves.
* Potential short evaluation of web server.

Goals for this week:

* Finish dissertation.