**Minutes**

19/01/23

* Went through what I had done since the previous meeting: advancements with FCN, Django server and RNNs.
* FCN model’s prediction was rounded up/down given a threshold of 0.5 and accuracy was calculated using an equality comparison with the label. Concern was that as there are more ordered labels, assuming all predictions were ordered could still result in a high accuracy.
  + Could consider weighting disordered predictions and labels more because they are unbalanced in the dataset.
  + Could implement Matthews correlation coefficient (MCC) as an accuracy measure – this is common for unbalanced predictions.
* Discussed BCE
  + <https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html>
  + The BCE follows , where a guaranteed prediction of an ordered amino acid would give q=0, giving us the log(1-p) probability.
  + Log probabilities are used due to multiplying many probabilities together.
  + I should make sure the BCE is behaving how I intend it to. As I have an array of labels, e.g., [0, 0, …, 0, 1, 1, …, 1, 0,…] it should consider all of the predictions compared to each label within the array not just the first index of the label for that protein sequence.
* Looked at validation data plotted at each epoch. Could plot accuracy/MCC instead of BCE loss to assess the model.
* Demo of Django server.
* Briefly discussed RNN from lab 6 of deep learning moodle.

Goals for this week:

* Create RNN model and use this with the project dataset so that I am sticking to the plan.
* Look at MCC, accuracy and how BCE is behaving with the FCN model.