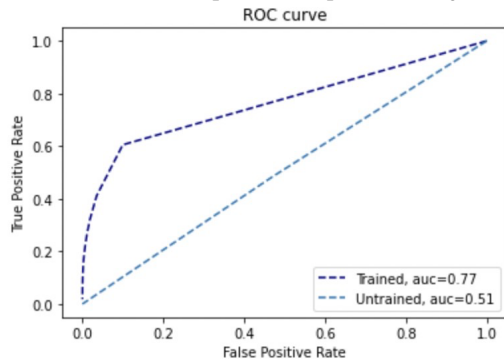# Problem 5 -Practical Report

5.1 ROC curve represents probability and AUC represents degree of separability. Model is good when AUC is near to

1. As we can see from figure AUC of untrained model is 0.51 which means model has no class separation capacity. Here AUC of trained model is 0.77 which means that there are 70% chance that model will be able to distinguish between positive and negative class. ROC curves visualize the trade-off between (True positive rate)TPR and (False positive rate)FPR. The curve is created by varying an internal decision threshold. Models with a curve closer to the upper left corner are better.

5.2 To compute Position weight matrix given below in table, I read .jasper file and converted them into NumPy array and then divided each value by sum of of column values.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0.10	0.18	0.31	0.06	0.01	0.81	0.04	0.12	0.93	0.01	0.37	0.06	0.01	0.06	0.11	0.41	0.09	0.13	0.44
0.32	0.16	0.05	0.88	0.99	0.01	0.58	0.47	0.01	0.00	0.00	0.01	0.00	0.01	0.81	0.01	0.53	0.35	0.20
0.08	0.45	0.49	0.02	0.00	0.07	0.37	0.05	0.04	0.99	0.62	0.55	0.98	0.85	0.01	0.56	0.34	0.08	0.29
0.50	0.20	0.15	0.04	0.00	0.10	0.01	0.36	0.02	0.00	0.01	0.37	0.01	0.08	0.07	0.02	0.04	0.44	0.06

Table 1: Position weight matrix-CTCF

5.3 Calculated normalized PWMs obtanied from 1st conv. layer and took max value over all filters resulting into 300 tensor.
5.4,5.5 Then counted the base-pair occurrences in the set of sequences from test set where result from activated filter is more than max over all filters divided by 2(i.e calculated from 5.3) using torch.functional.unfold . There after calculated pearson correlation of 300 PWMS with CTCF PWM and obtained most correlated PWM with 0.76 correlation at index 180 from test set where model was trained with learning rate 0.002. Whereas obtained most correlated PWM at index 277 with correlation 0.67 for model trained at learning rate 0.002(default), which is a bit of surprise since most correlated PWM's changes when model changes. As we can see from below image, position weight of A, G is similar in CTCF and PWM with correlation 0.76.