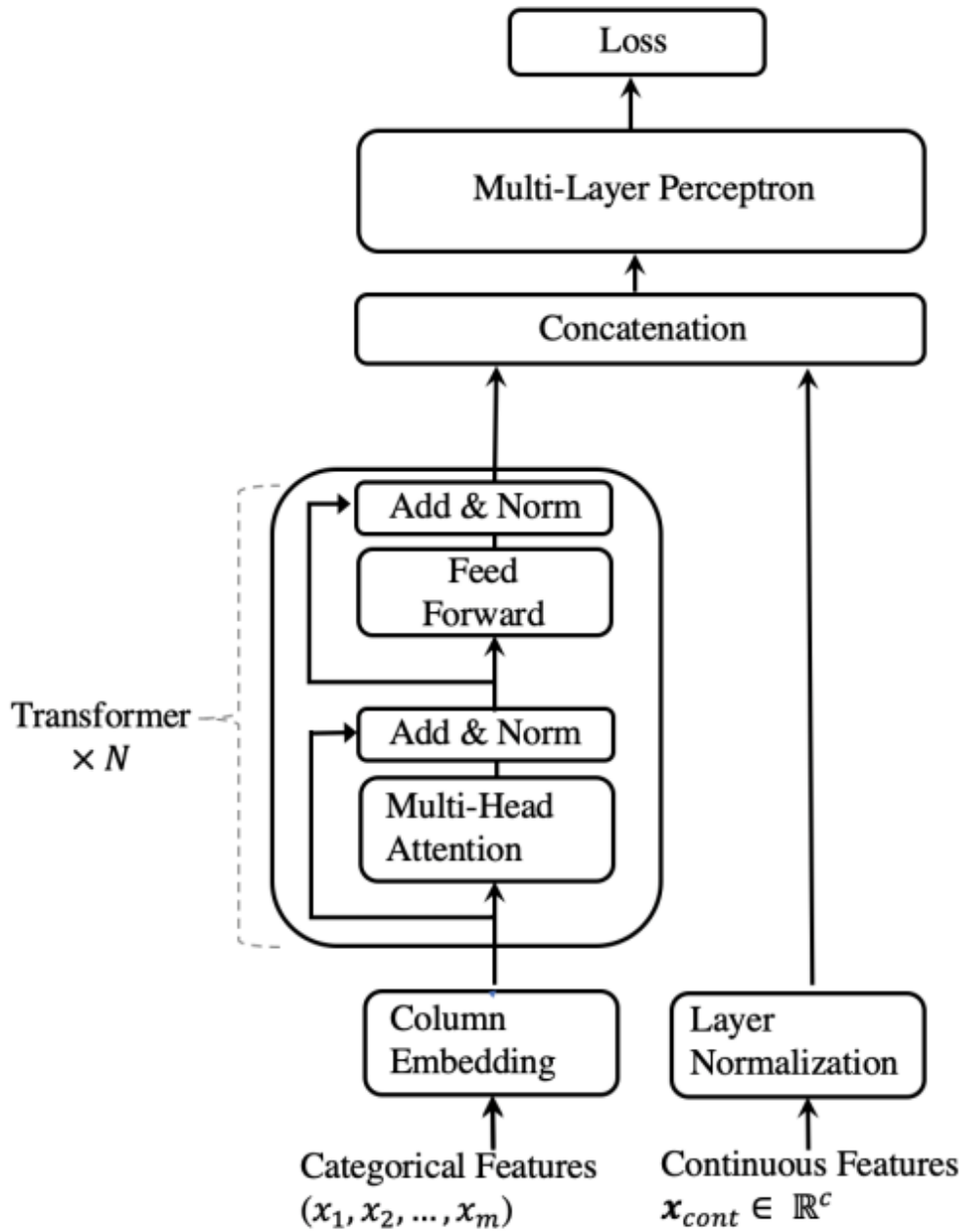
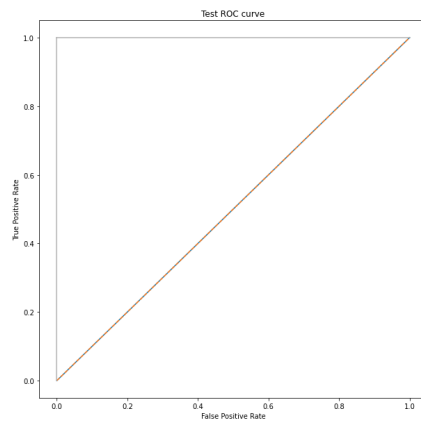
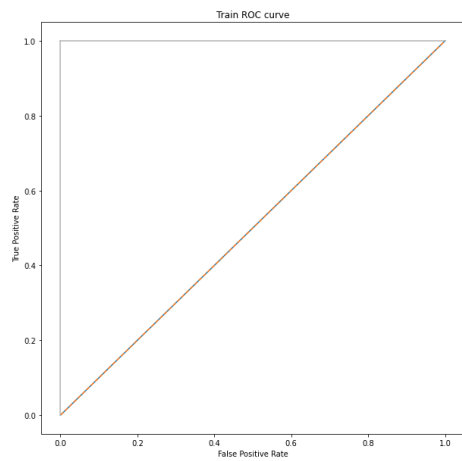


Architecture:

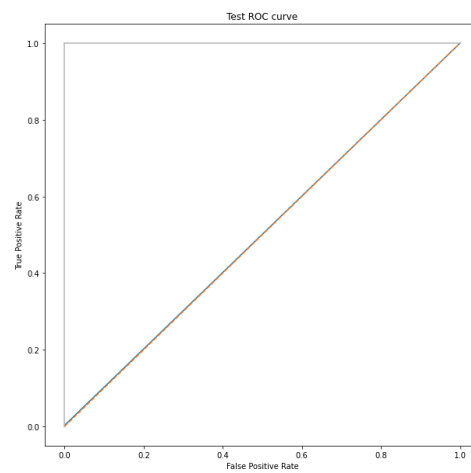
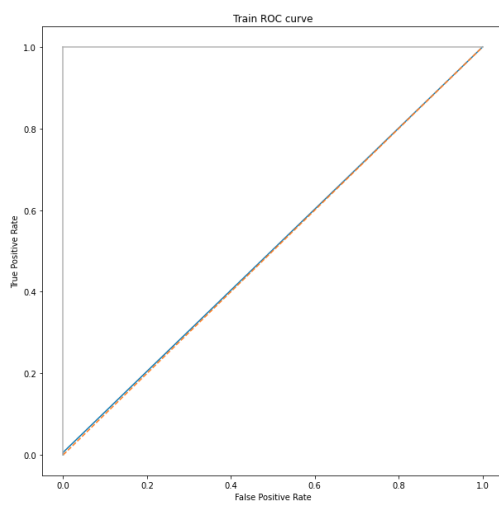


Datasets: <https://www.kaggle.com/shrutimechlearn/churn-modelling>

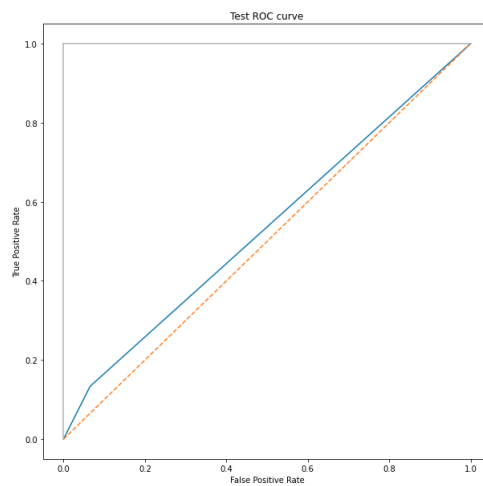
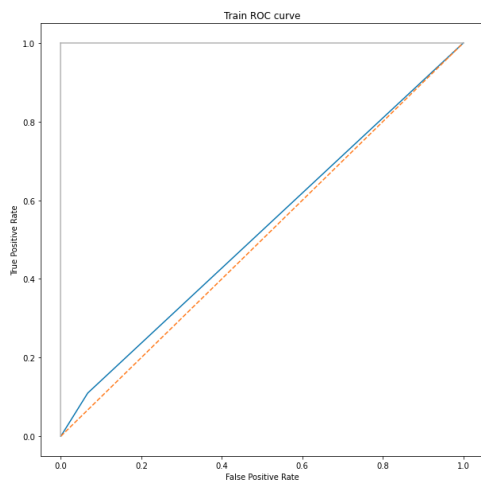
Heads = 1, depth = 6, attn_dropout=0.1, ff_dropout=0.1



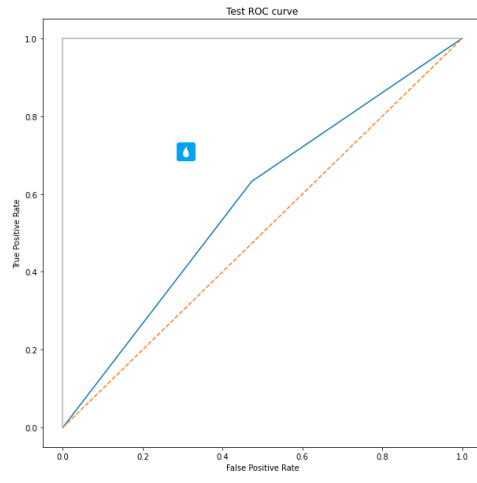
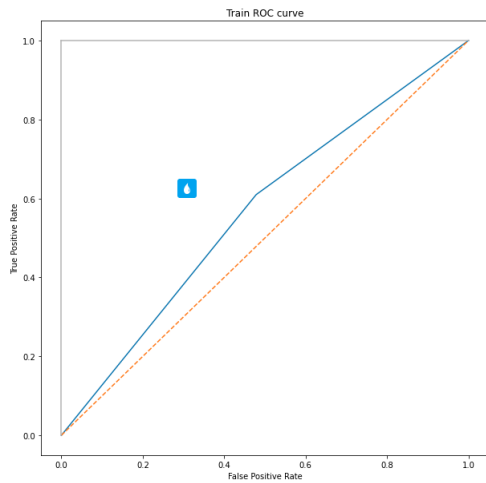
Heads = 2, depth = 6, attn_dropout=0.1, ff_dropout=0.1



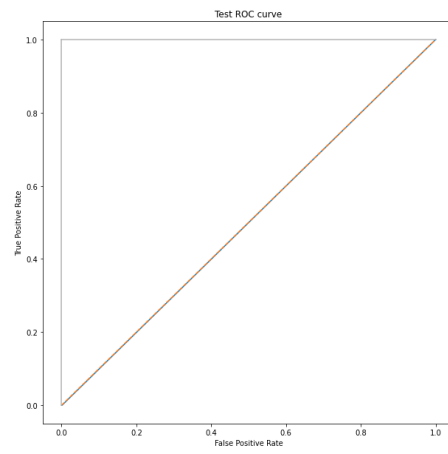
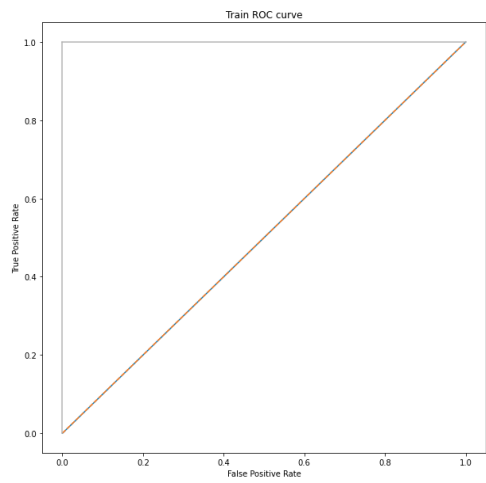
Heads = 4, depth = 6, attn_dropout=0.1, ff_dropout=0.1



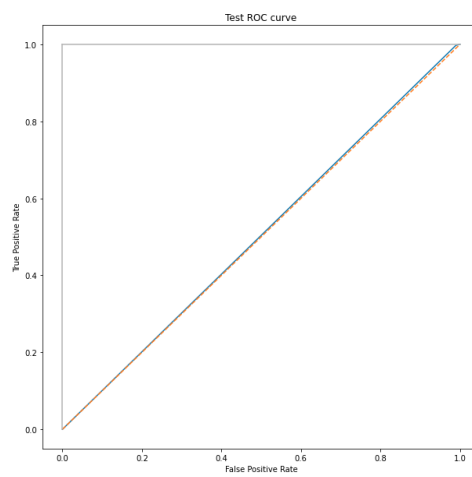
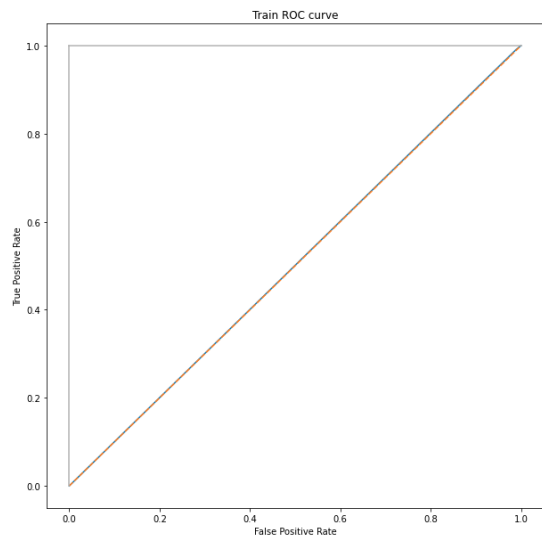
Heads = 8, depth = 6, attn_dropout=0.1, ff_dropout=0.1



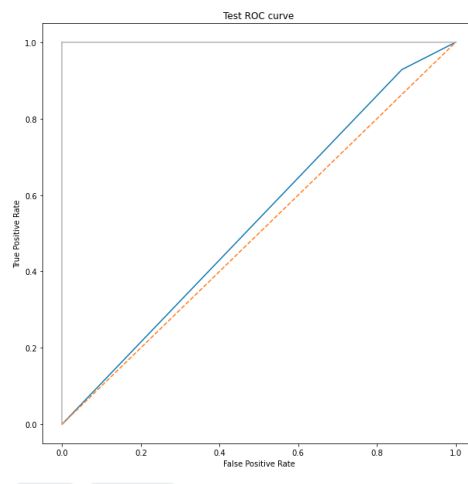
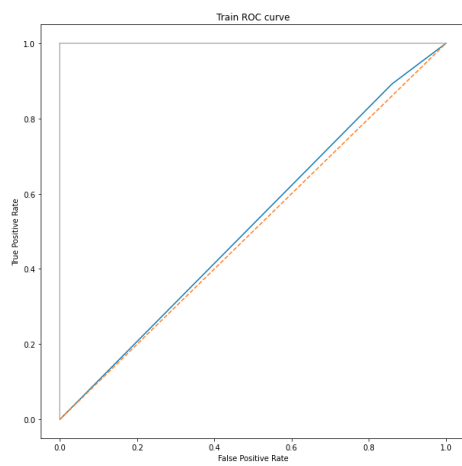
Heads = 16, depth = 6, attn_dropout=0.1, ff_dropout=0.1



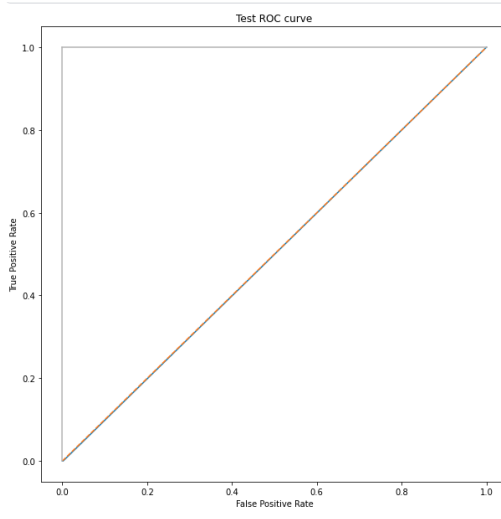
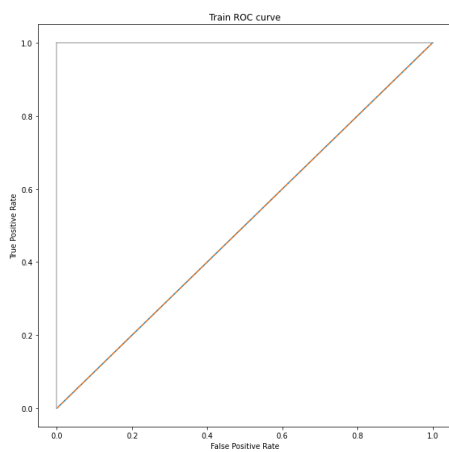
Heads = 32, depth = 6, attn_dropout=0.1, ff_dropout=0.1



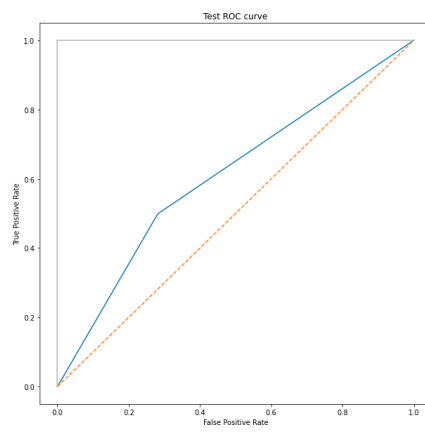
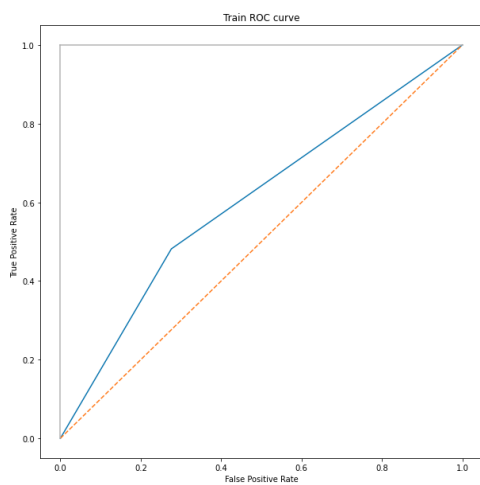
Heads = 8, depth = 6, attn_dropout=0.0, ff_dropout=0.



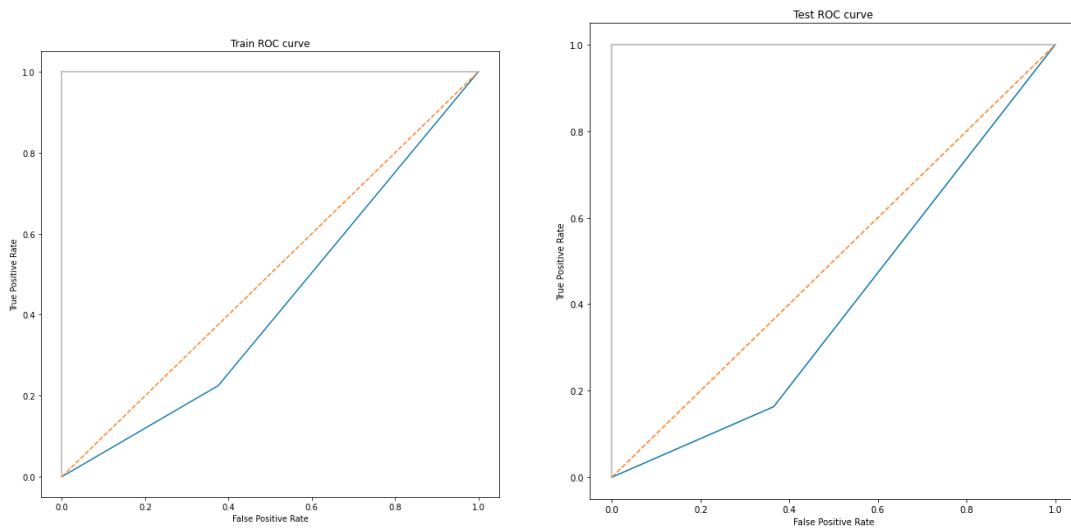
Heads = 8, depth = 6, attn_dropout=0.5, ff_dropout=0.5



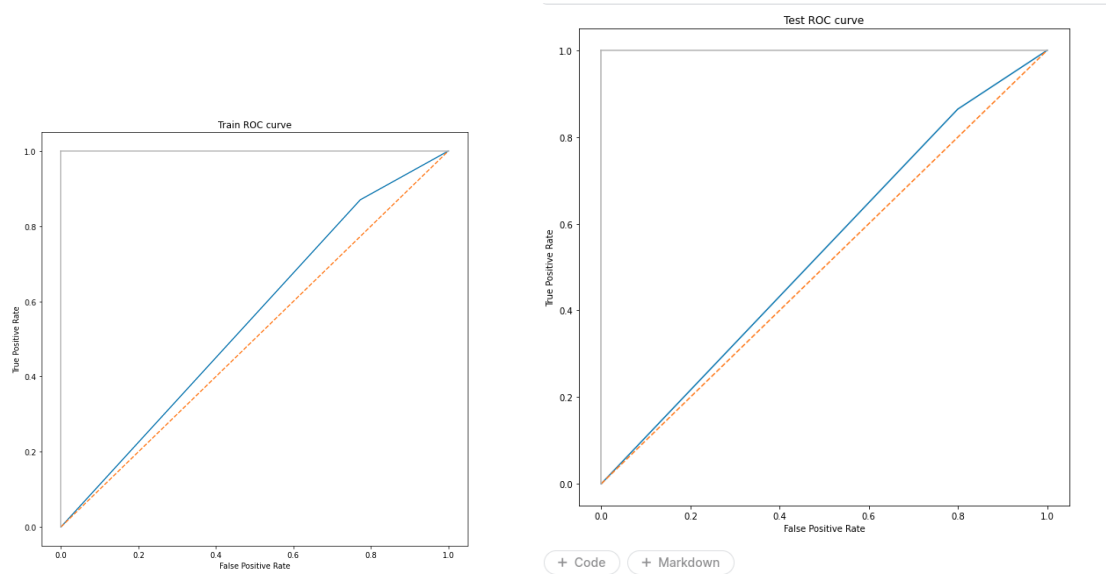
Heads = 8, depth = 2, attn_dropout=0.1, ff_dropout=0.1



Heads = 8, depth = 5, attn_dropout=0.1, ff_dropout=0.1

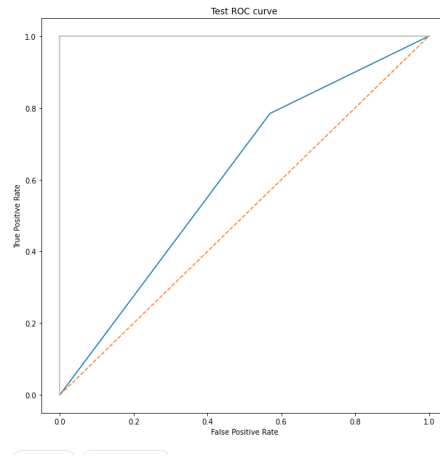
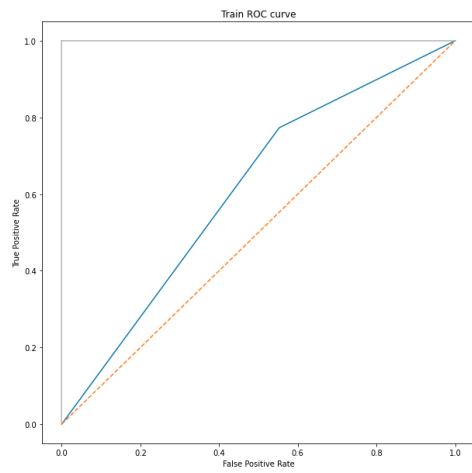


Heads = 8, depth = 6, attn_dropout=0.1, ff_dropout=0.1,
mlp_hidden = (5,3)

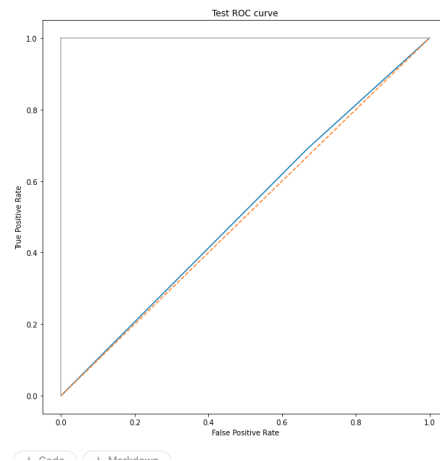
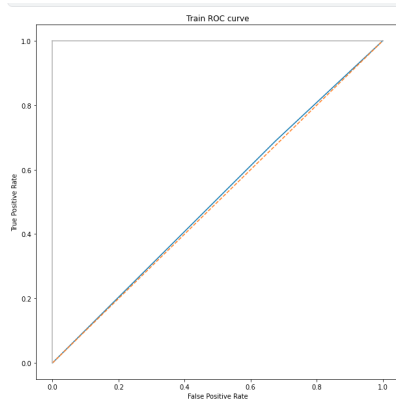


Dataset: <https://www.kaggle.com/lodetomasi1995/income-classification>

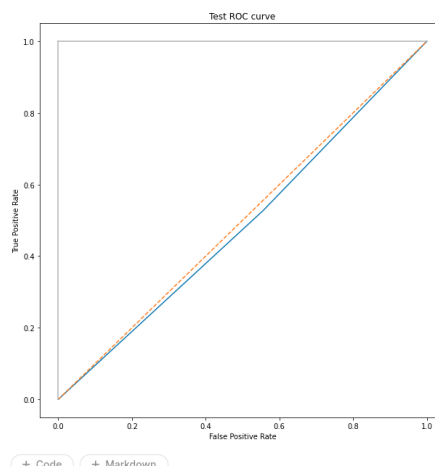
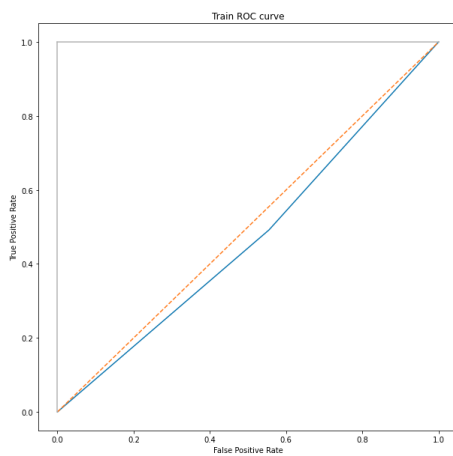
Heads = 8, depth = 6, attn_dropout=0.1, ff_dropout=0.1



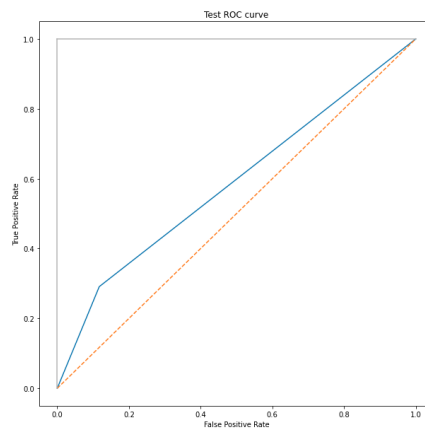
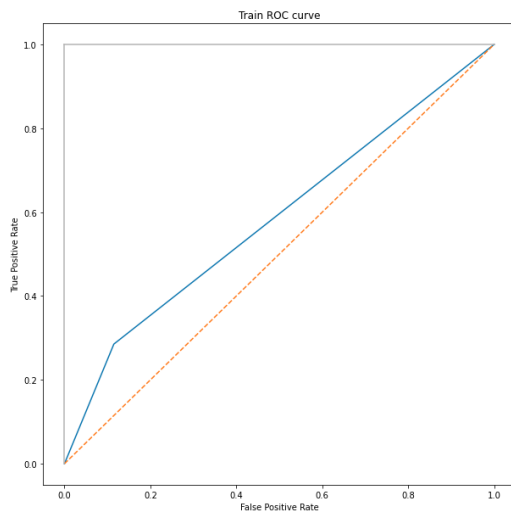
Heads = 1, depth = 6, attn_dropout=0.1, ff_dropout=0.1



Heads = 2, depth = 6, attn_dropout=0.1, ff_dropout=0.1



Heads = 4, depth = 6, attn_dropout=0.1, ff_dropout=0.1



Heads = 8, depth = 6, attn_dropout=0.0, ff_dropout=0.

