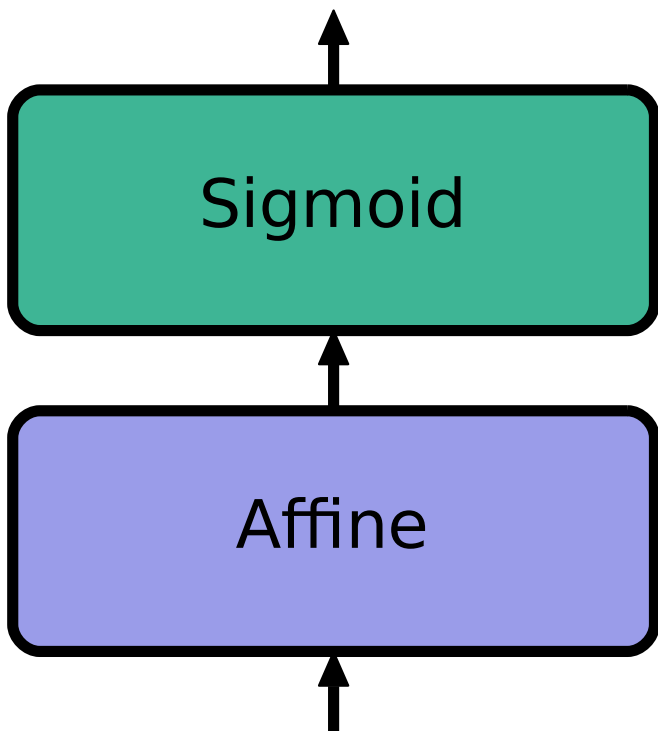


Classification Head

Output (N, L, 1)



Input (N, L, D_model)