Original params:

episodes = 100\_000 # Total number of episodes

warmup\_steps = 170\_000 # Amount of warmup steps to collect data with random policy

memory\_length = warmup\_steps # Max length of the Replay Memory

batch\_size = 32

collect\_steps\_per\_episode = 2000

collect\_every = 500

target\_update\_period = 800 # Period to overwrite the target Q-network with the default Q-network

target\_update\_tau = 1 # Soften the target model update

n\_step\_update = 1

# reg = regularizers.l2(0.01)

layers = [Dense(256, activation="relu"), Dropout(0.2),

Dense(256, activation="relu"), Dropout(0.2),

Dense(256, activation="relu"), Dropout(0.2),

Dense(2, activation=None)] # No activation, pure Q-values

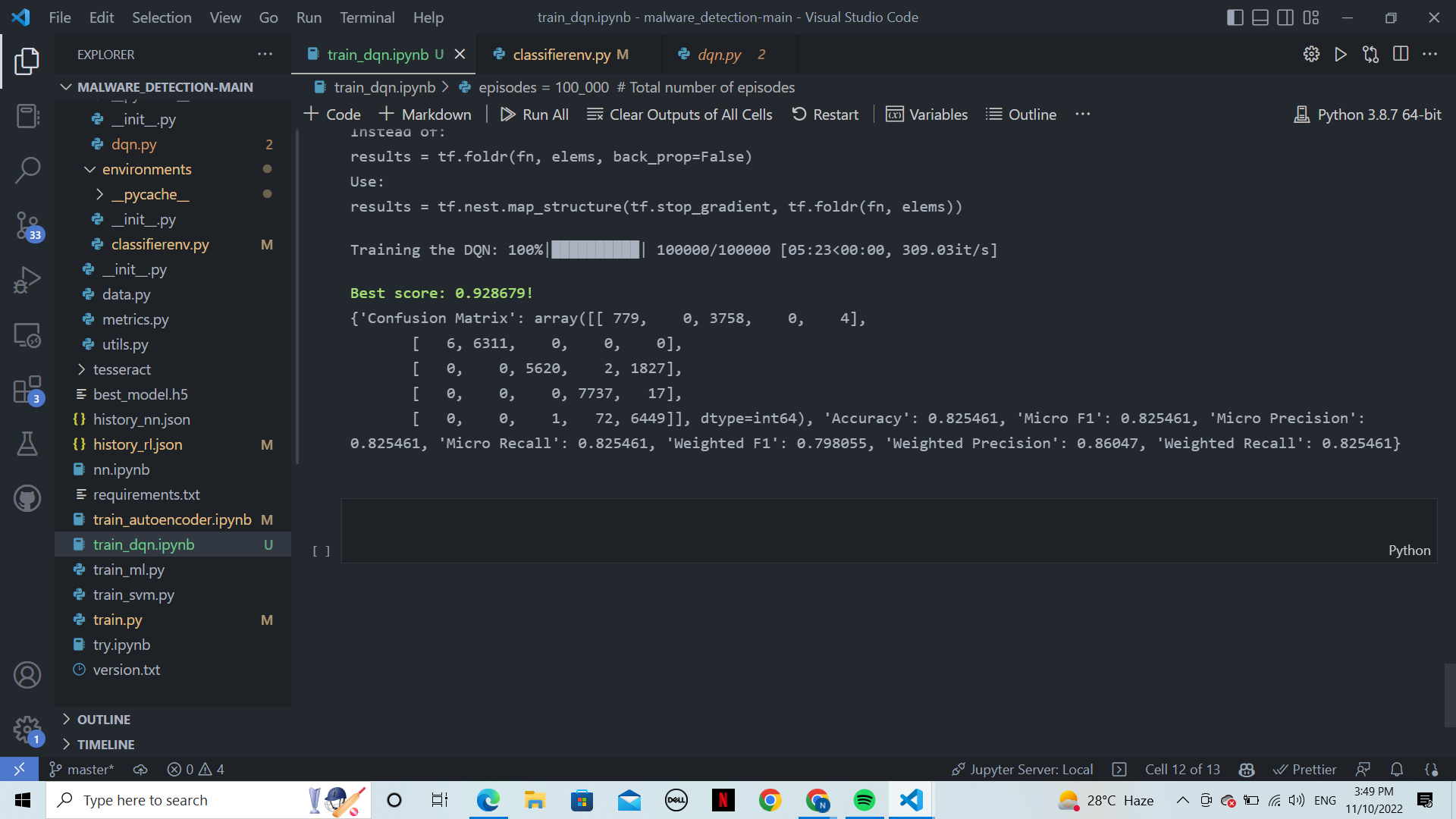
learning\_rate = 0.00025 # Learning rate

gamma = 0.1 # Discount factor

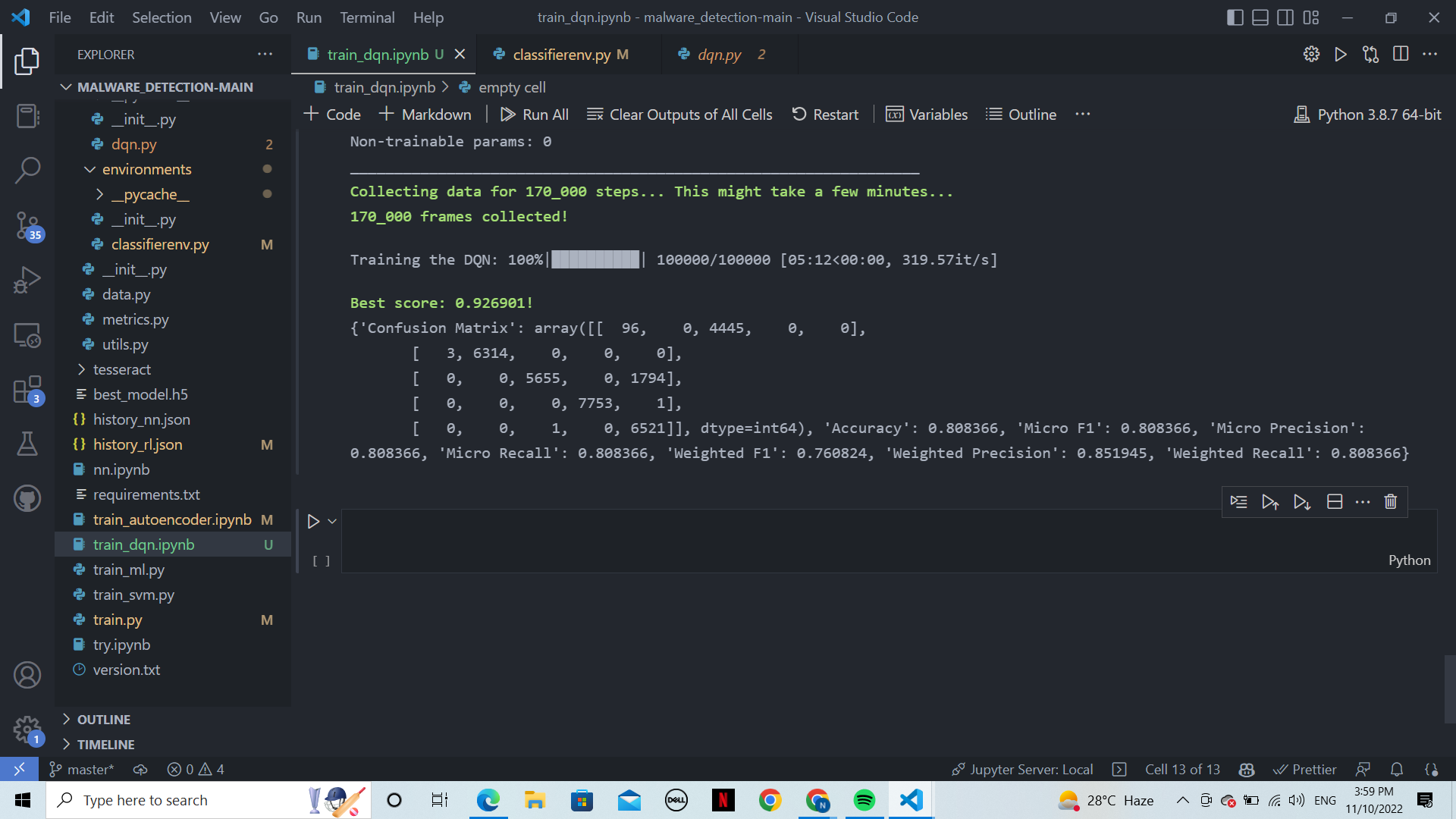
min\_epsilon = 0.5 # Minimal and final chance of choosing random action

decay\_episodes = episodes // 10 # Number of episodes to decay from 1.0 to `min\_epsilon``

1. Original params



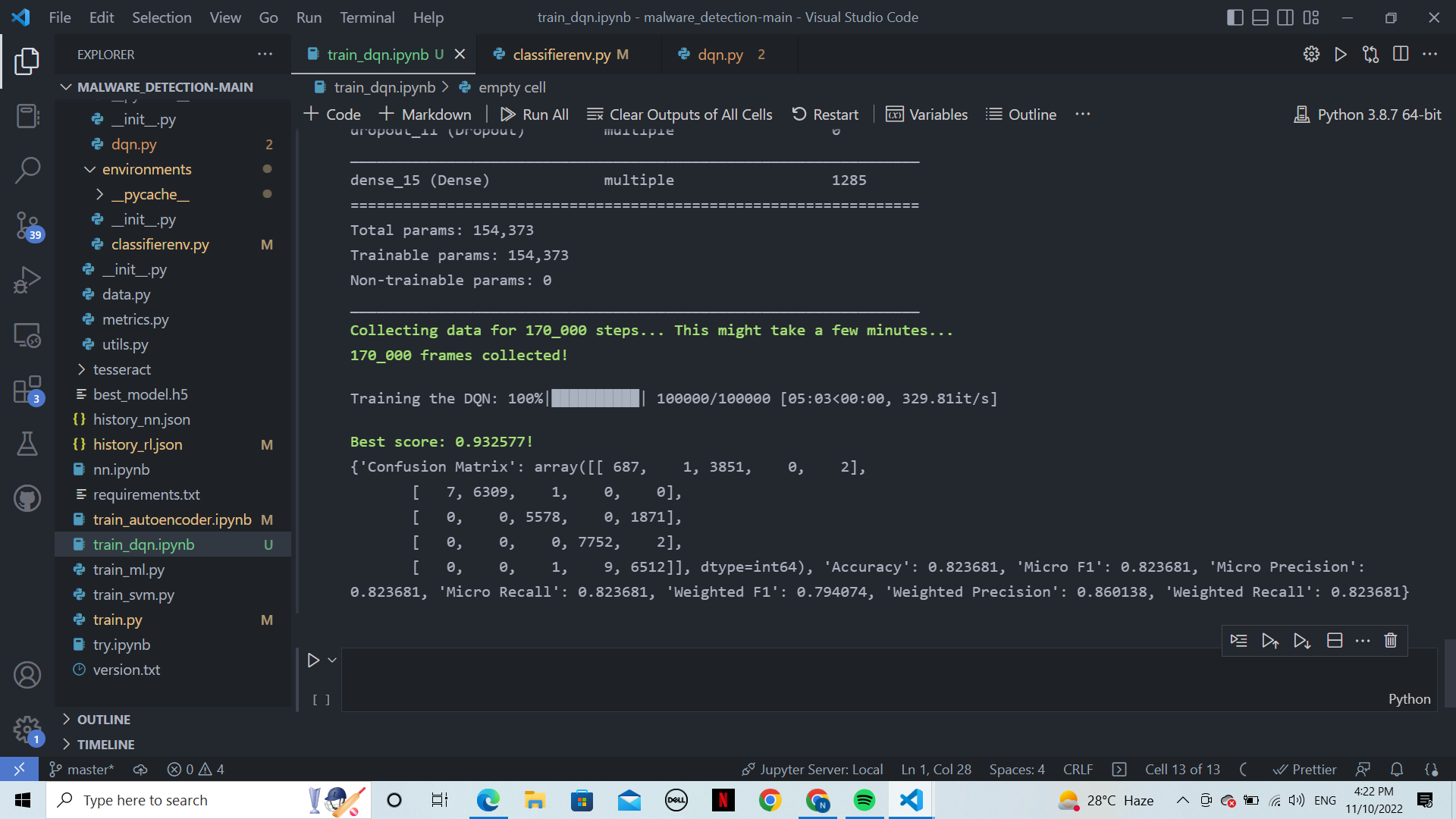
1. Activation: tanh



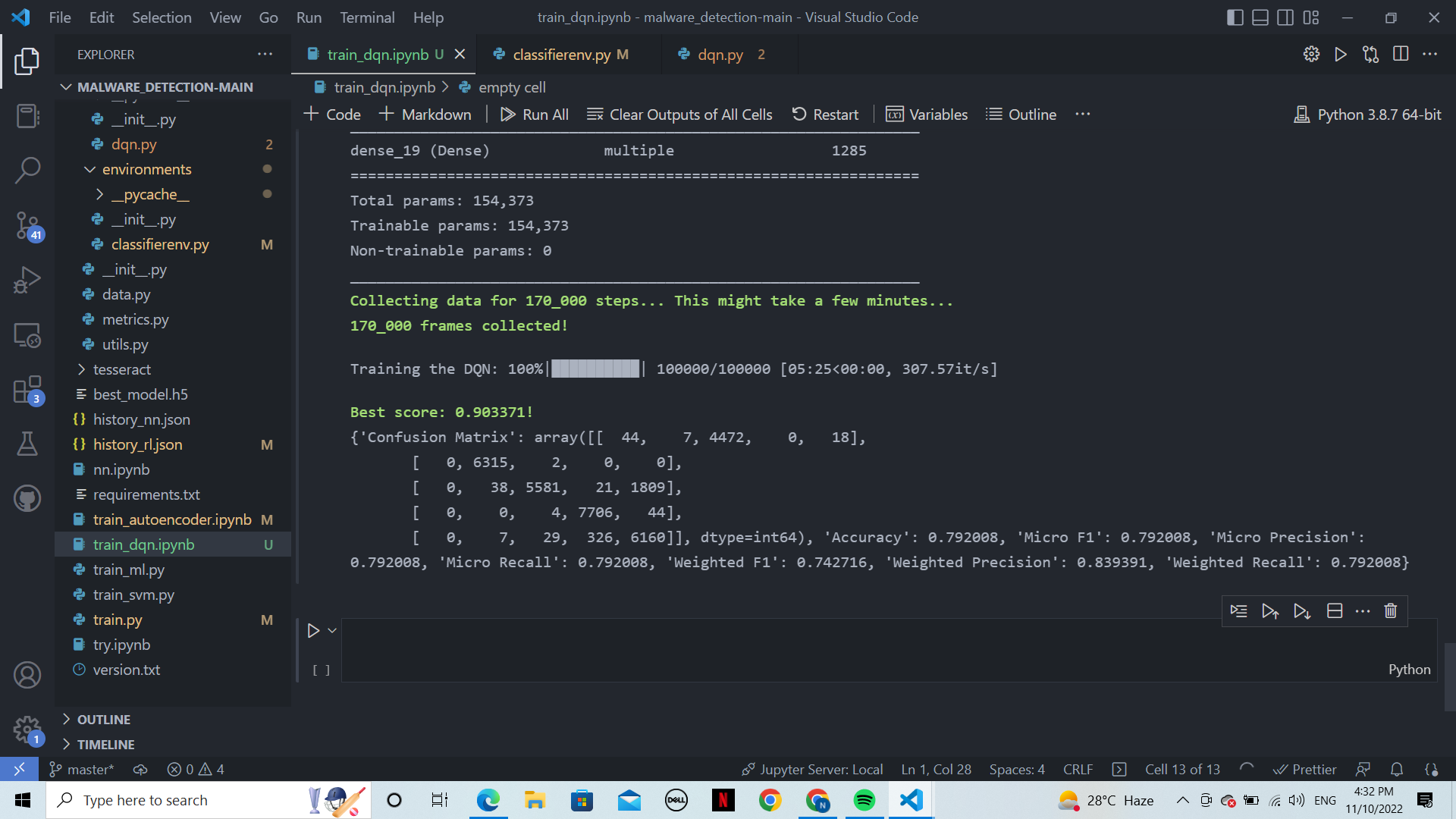
1. activation : sigmoid



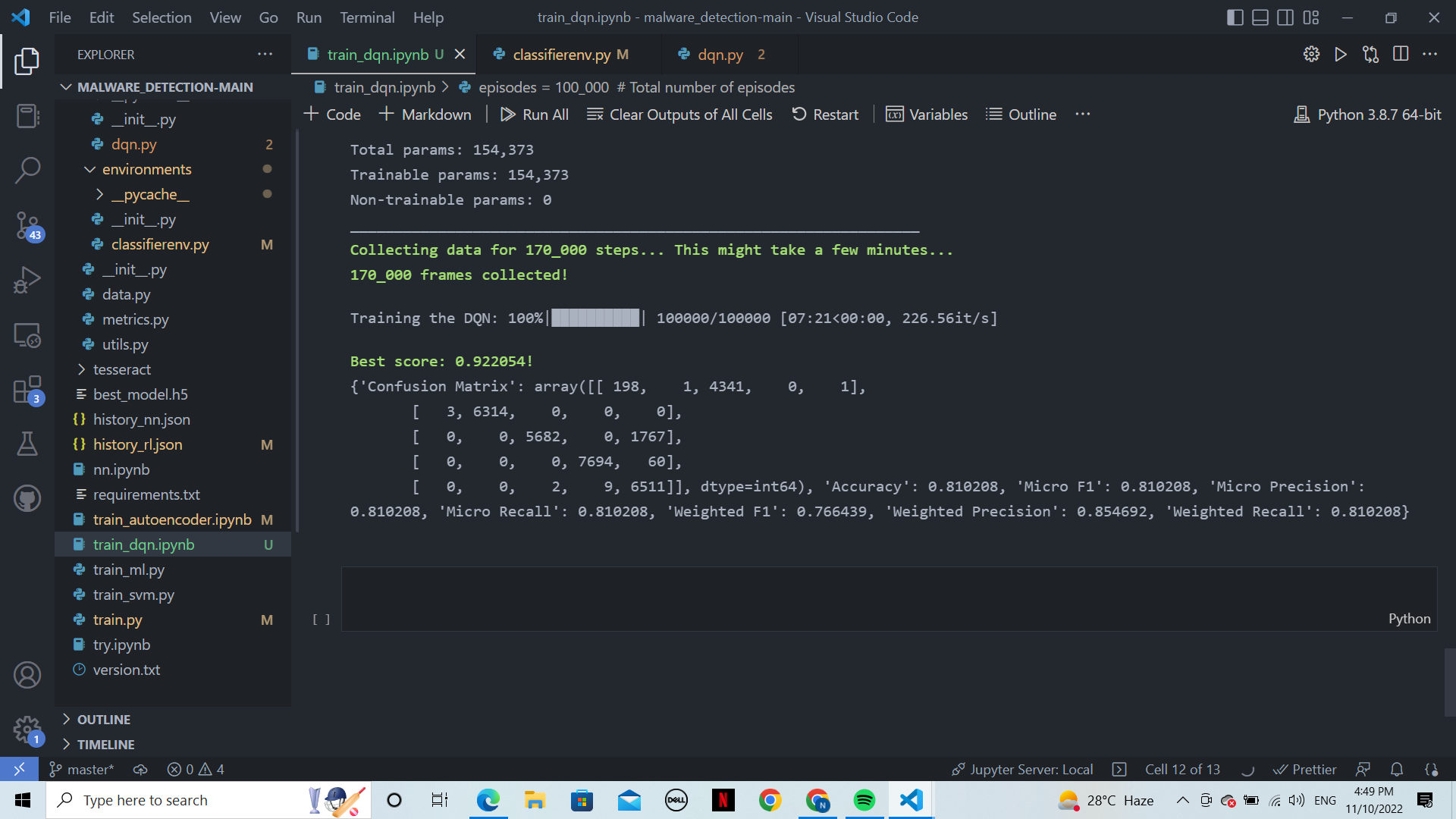
1. Dropout: 0.3



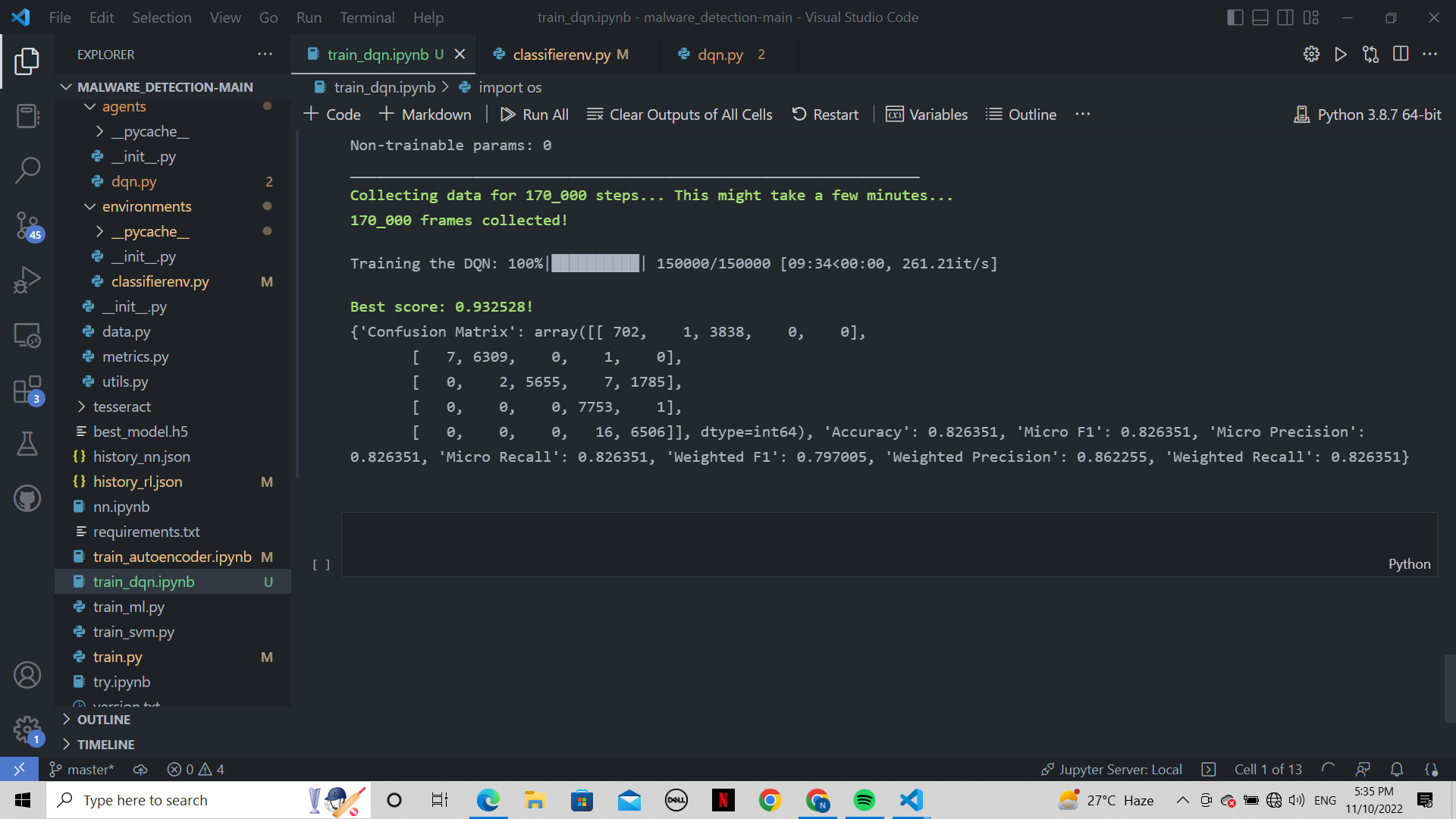
1. Dropout: 0.1



1. Learning rate: 0.000025 (initial/10)

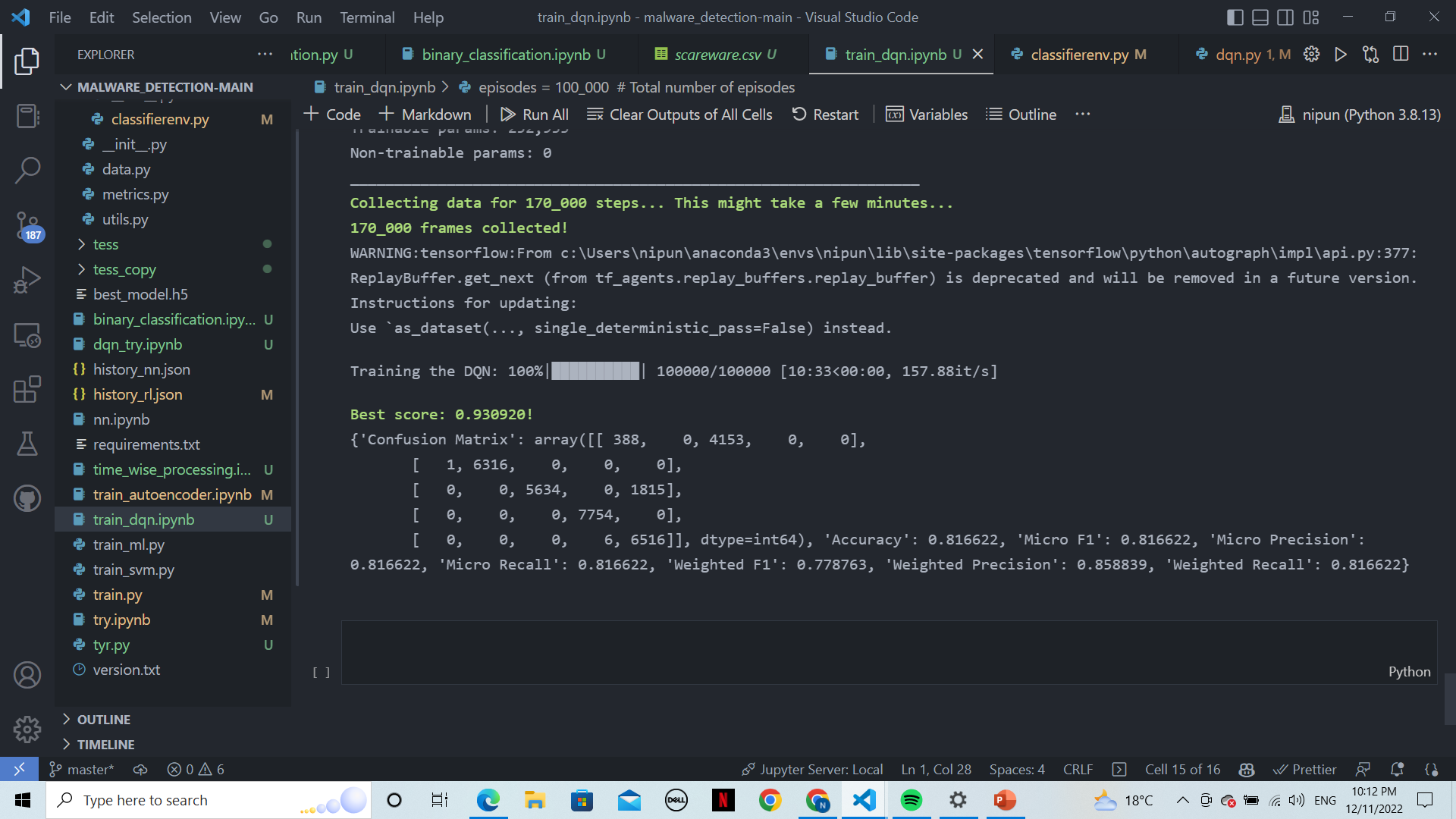


1. Learning rate: 0.0025 (initial \* 10)

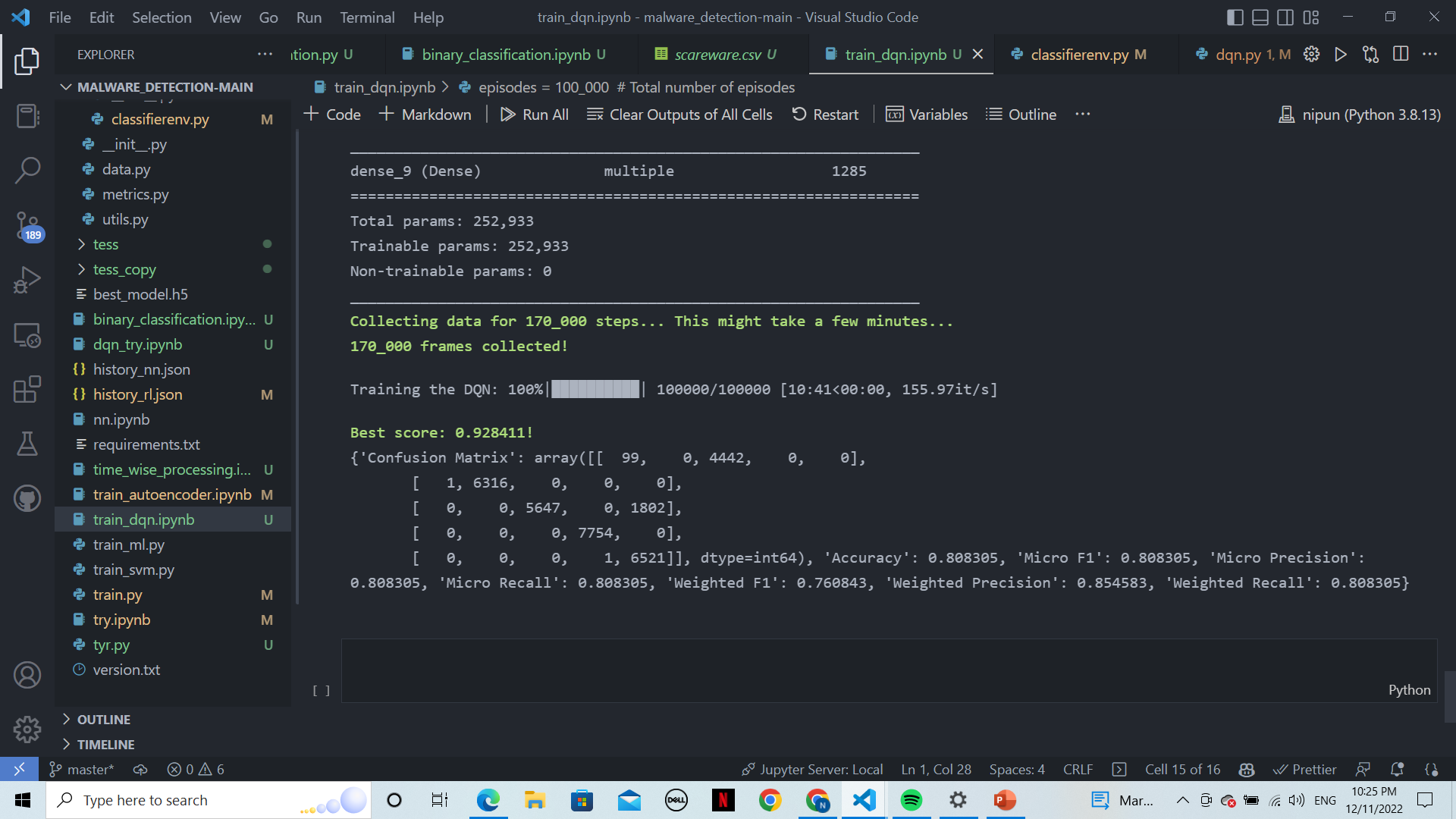


DQN + LSTM:

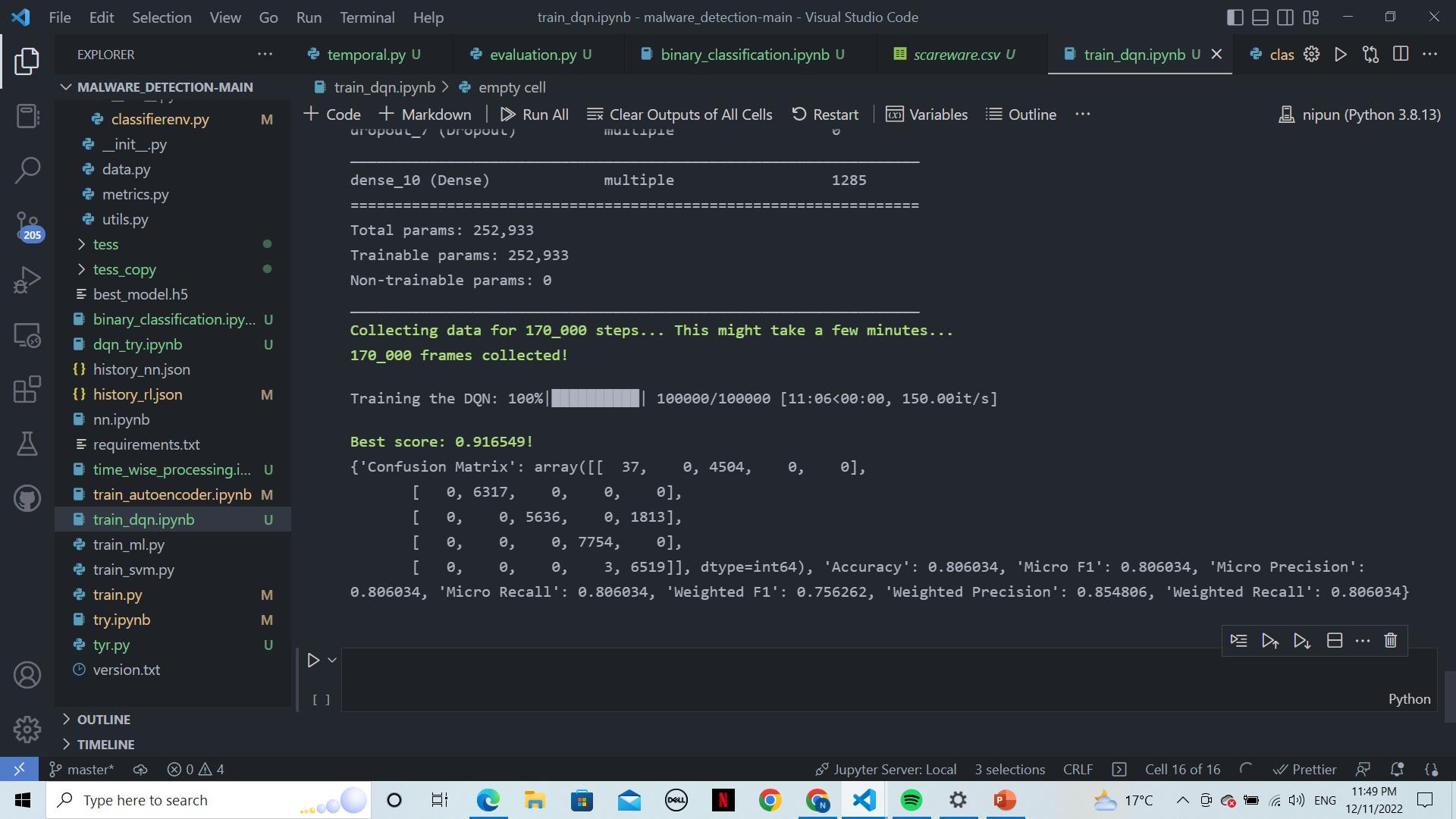
1. Original params:



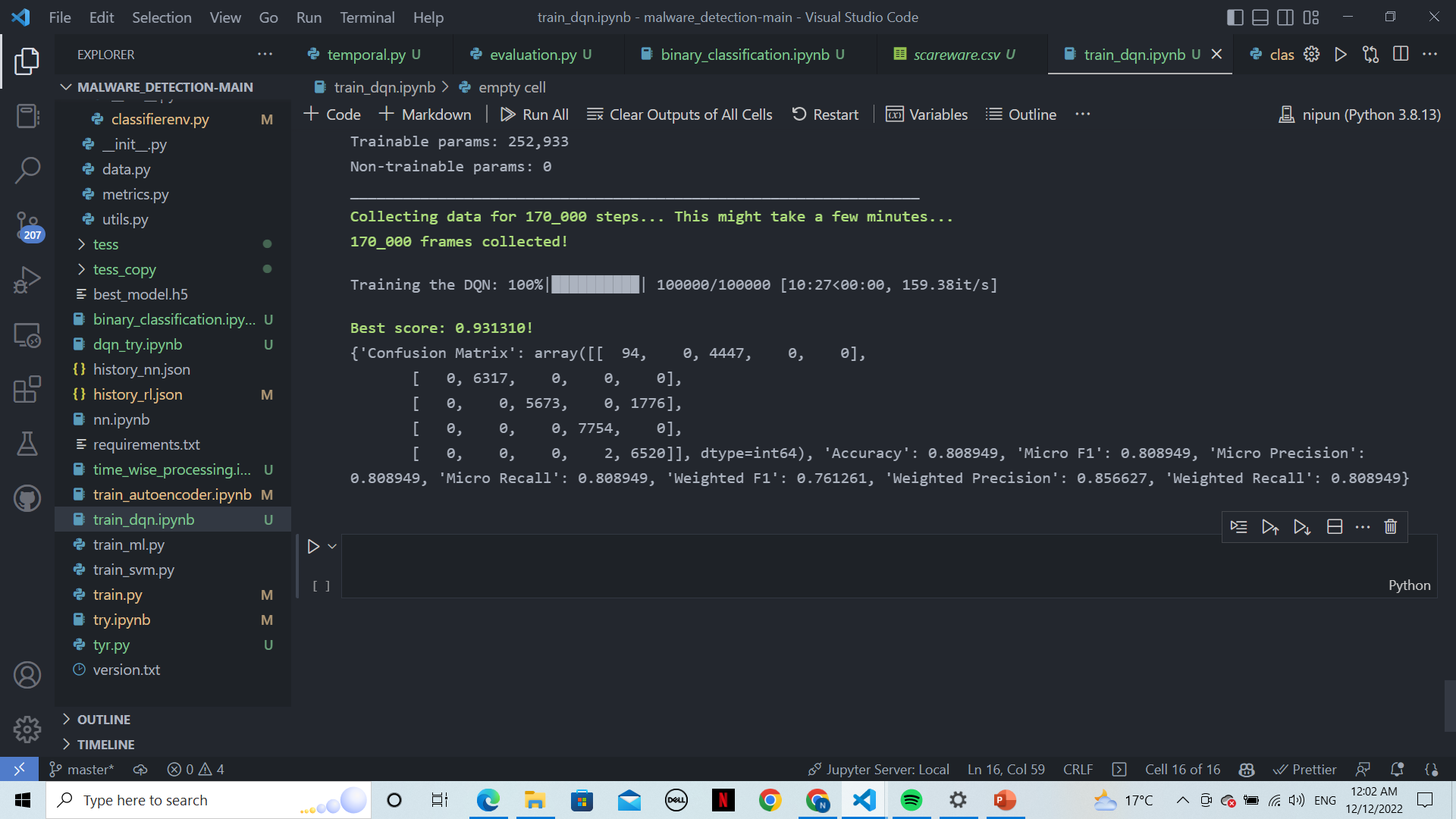
1. Activation: tanh



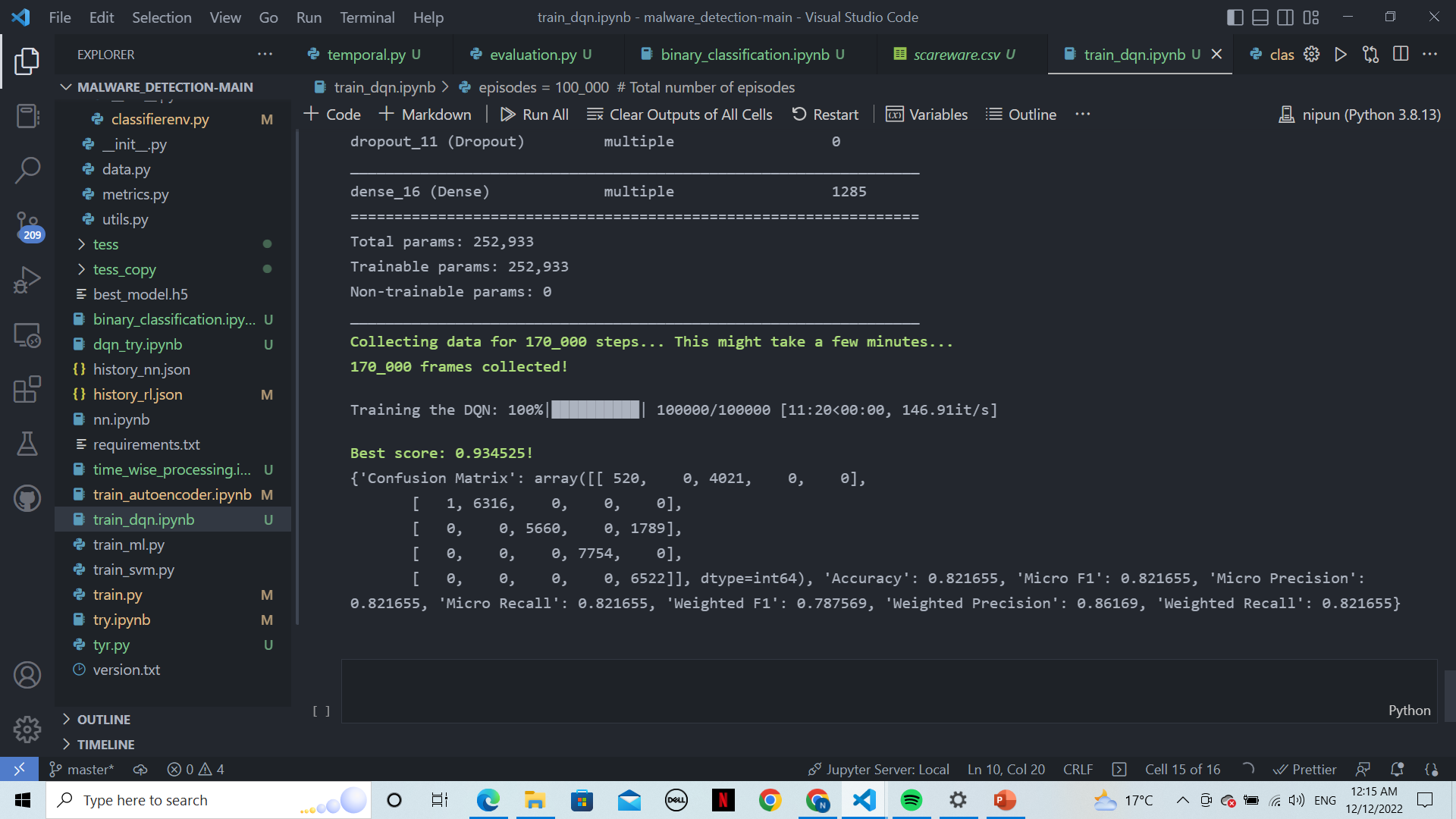
1. activation : sigmoid



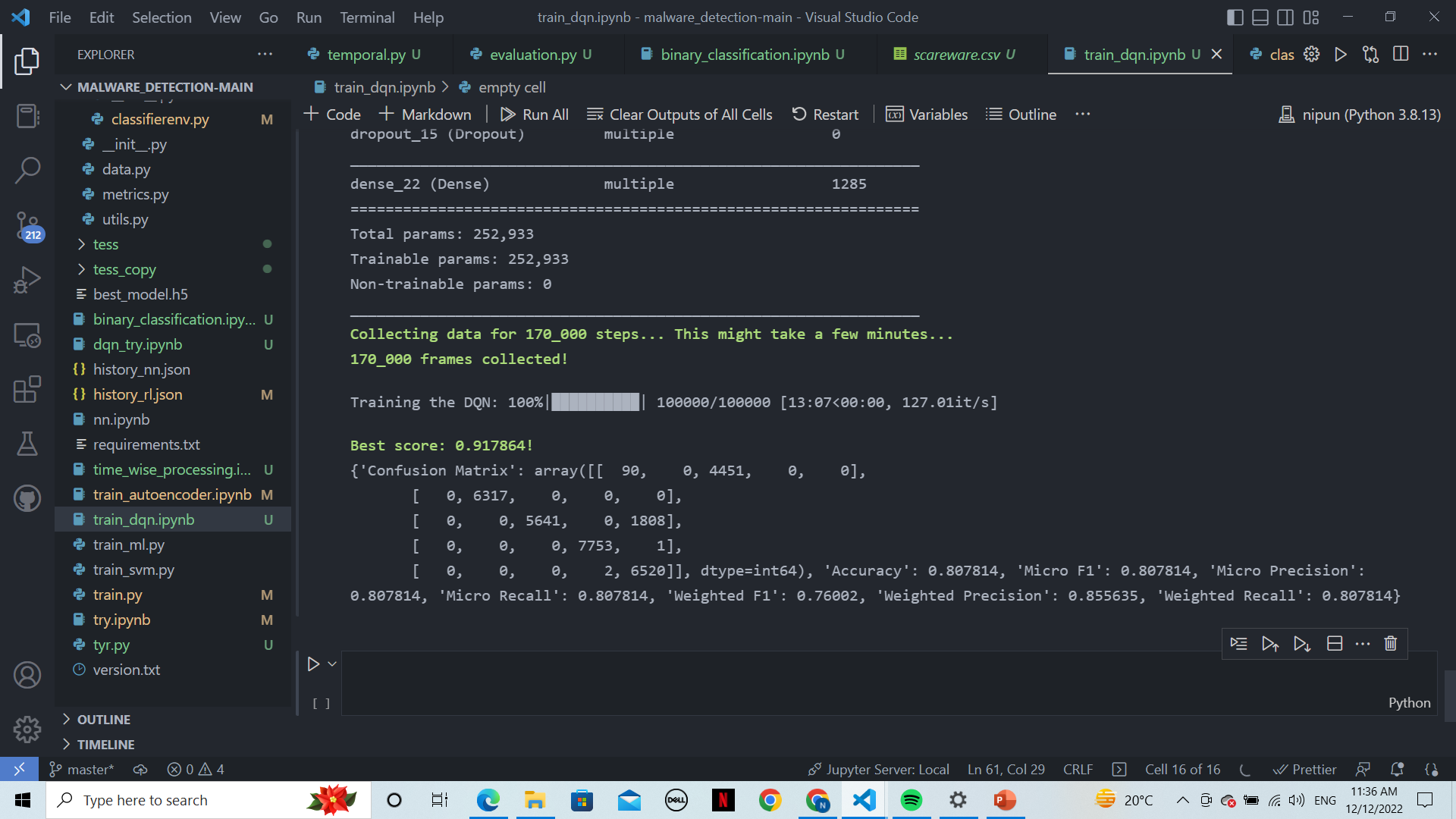
1. Dropout: 0.3



1. Dropout: 0.1



1. Learning rate: 0.000025 (initial /10):



1. Learning rate: 0.0025 (initial \* 10)

