3.5 Plot the average normalized training error and the average normalized test error over 10 trials vs r. Comment on the results, in relation to the results from part (3.2) where you explored different ℓ2 regularization coefficients. Can you provide an explanation for the behavior seen in this plot?

Normalized train error (SGD):
[0.5671845217864043, 0.5682738667530166, 0.5681697019072296, 0.5691155644850522, 0.6838982121877393, 0.9337740263739082, 1.232233815248266]

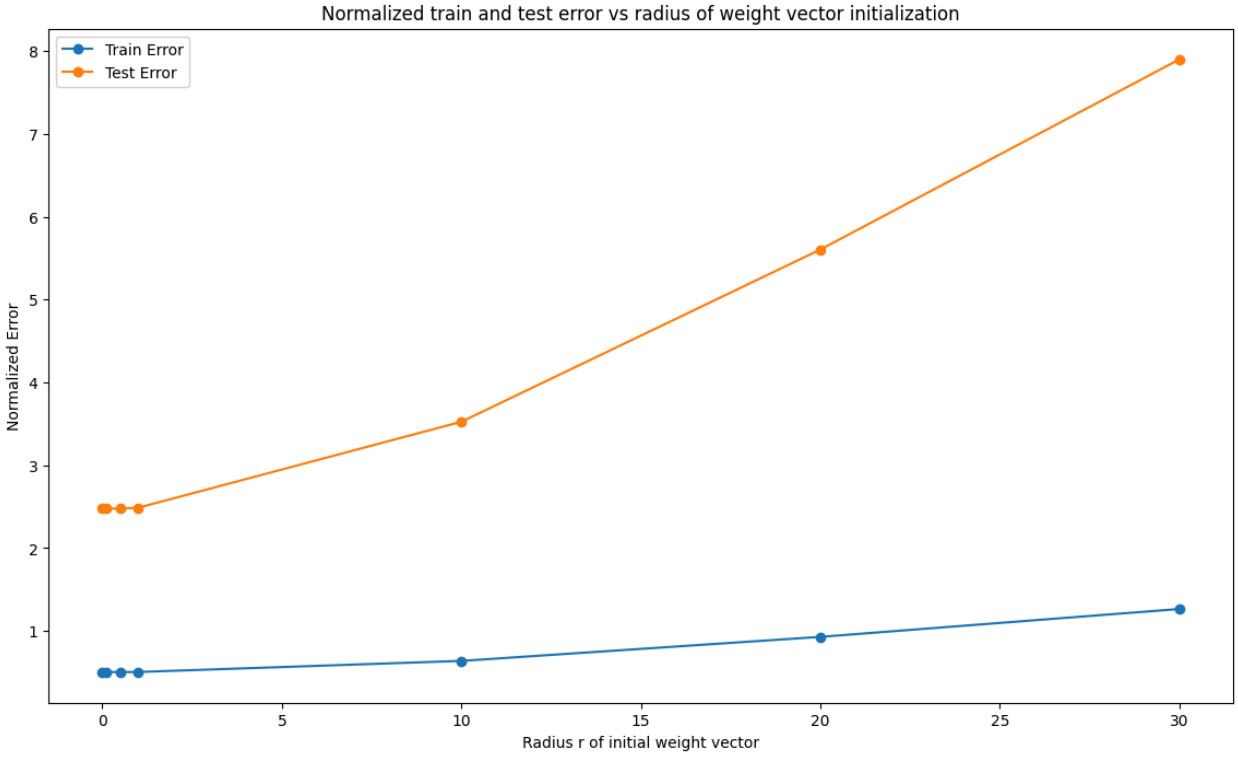
Normalized test error (SGD):
[3.7431243645040784, 3.7425923070178024, 3.7470914158790825, 3.7528951628033456, 4.8929737225077945, 7.283384055505208, 10.082784674547897]

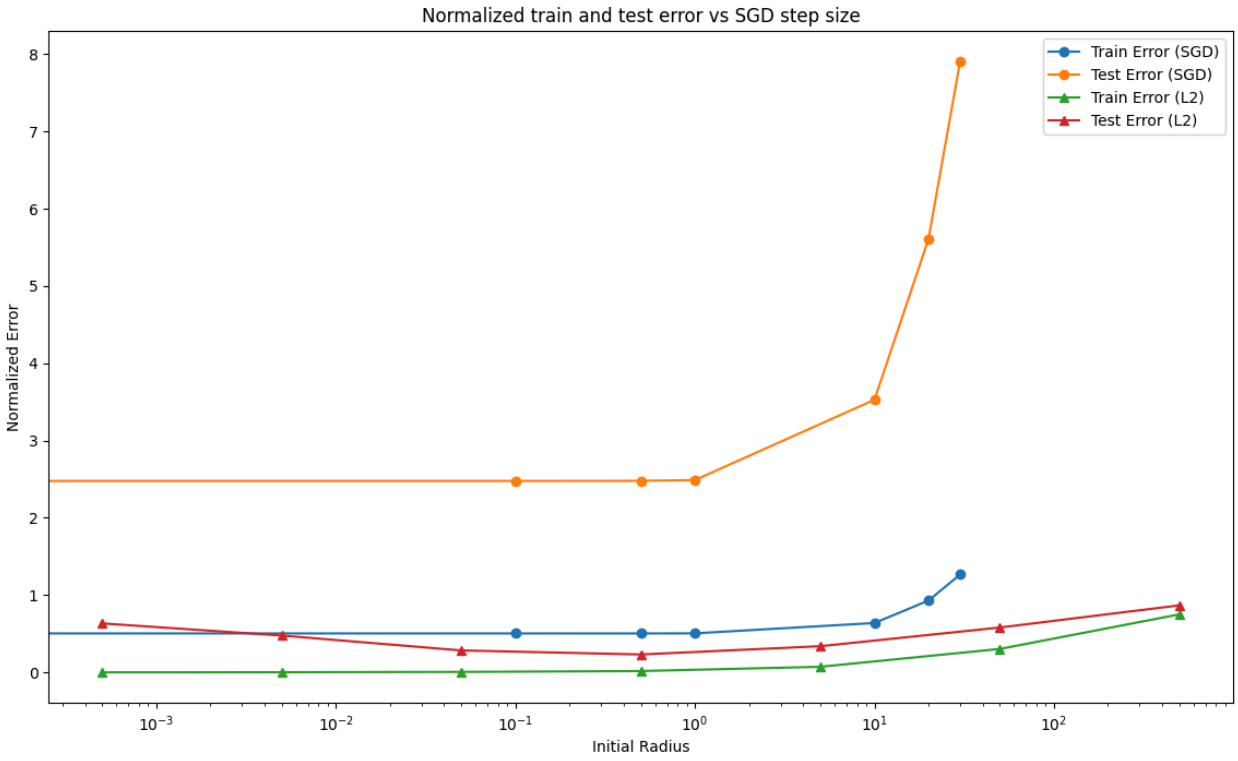
A larger radius leads to a higher normalized error. While the train error increases slowly when the radius is larger, the test error increases much more than the train error.

We also notice that for a larger radius, the norm is larger as well. Given the observation in 3.4, a smaller norm leads to lower test error with a better generalization.

Looking at this graph and the one obtained in the earlier section, the patterns of training errors and test errors can be compared.

From problem 3.2, the best lambda with the lowest test error is 0.5. In problem 3.5, the test error is also close to the lowest error for radius 0.5.




Normalized train and test error vs SGD step size
Train Error (SGD)
Test Error (SGD)
Train Error (L2)
Test Error (L2)
Normalized Error
Initial Radius
0
1
2
3
4
5
6
7
8
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}