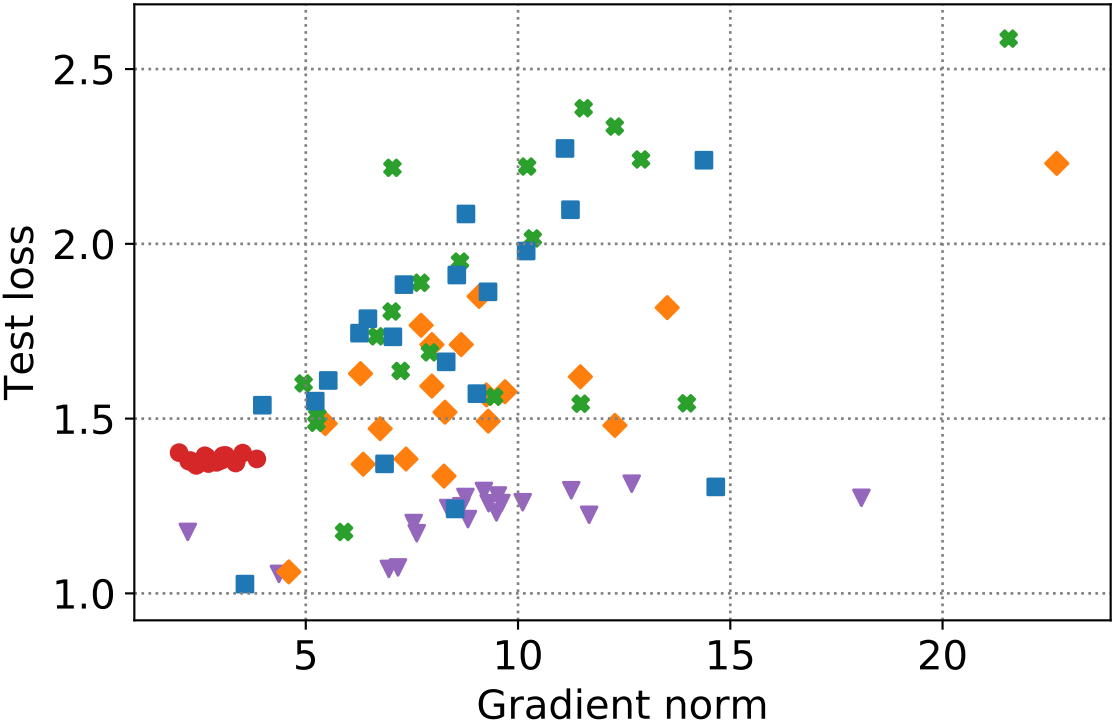
Test loss
2.5
2.0
1.5
1.0
5
10
15
20
Gradient norm