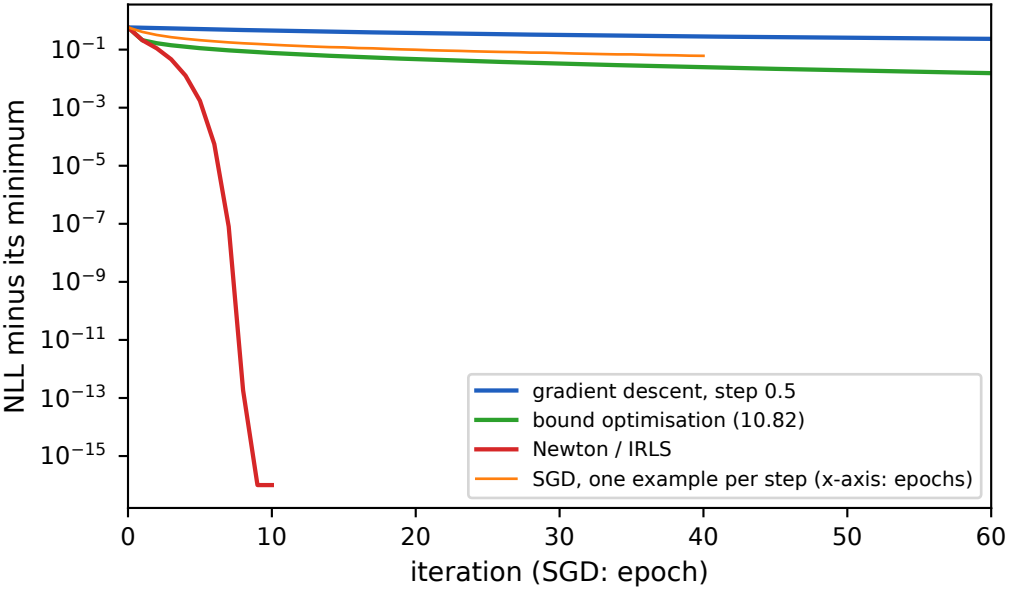

NLL minus its minimum
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
10^{-15}
0
10
20
30
40
50
60
iteration (SGD: epoch)
gradient descent, step 0.5
bound optimisation (10.82)
Newton / IRLS
SGD, one example per step (x-axis: epochs)