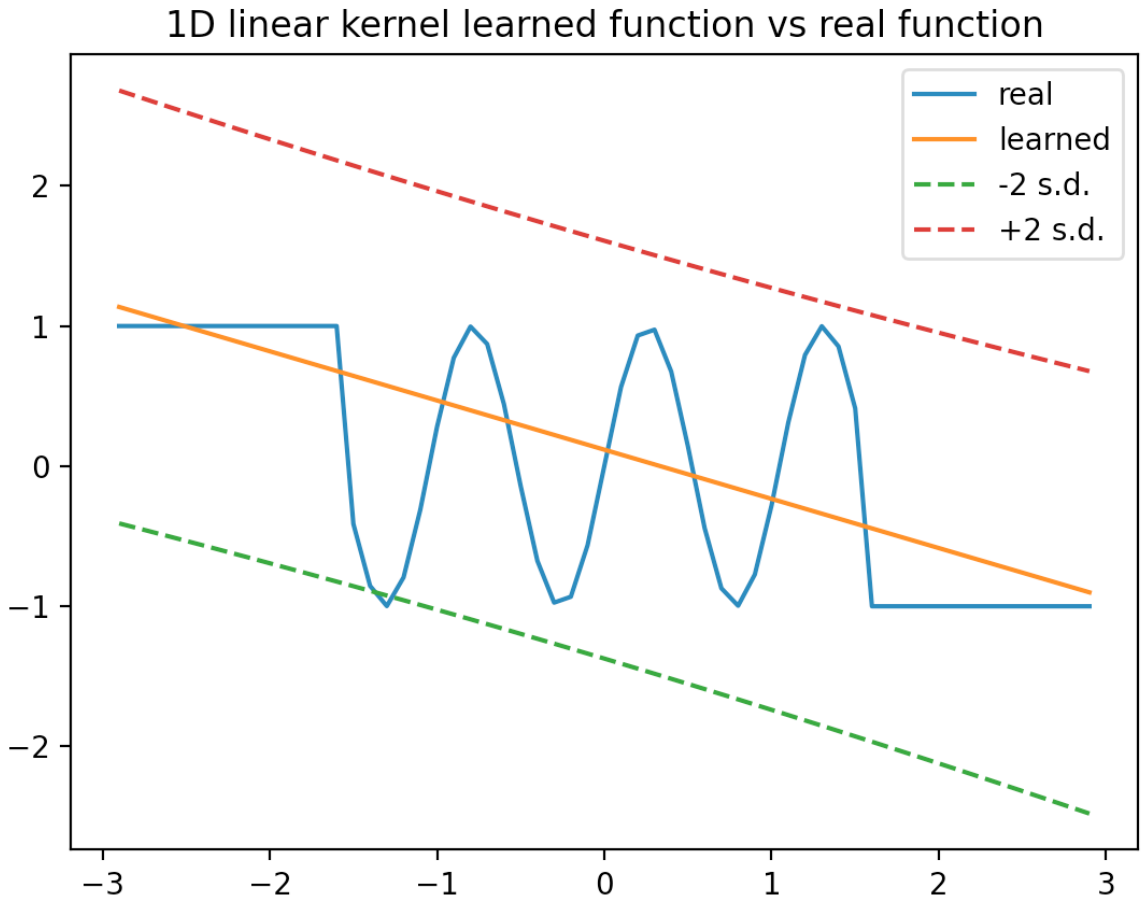
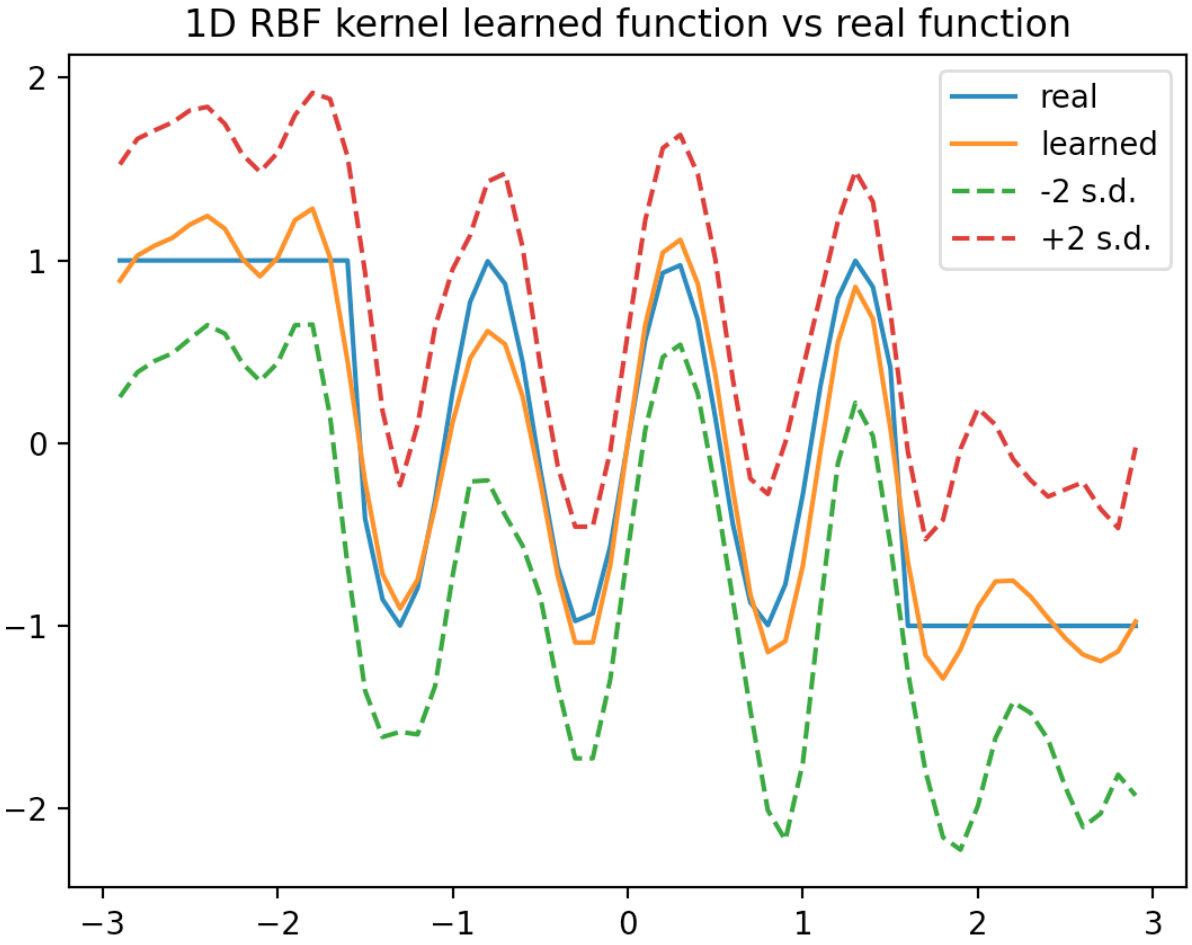
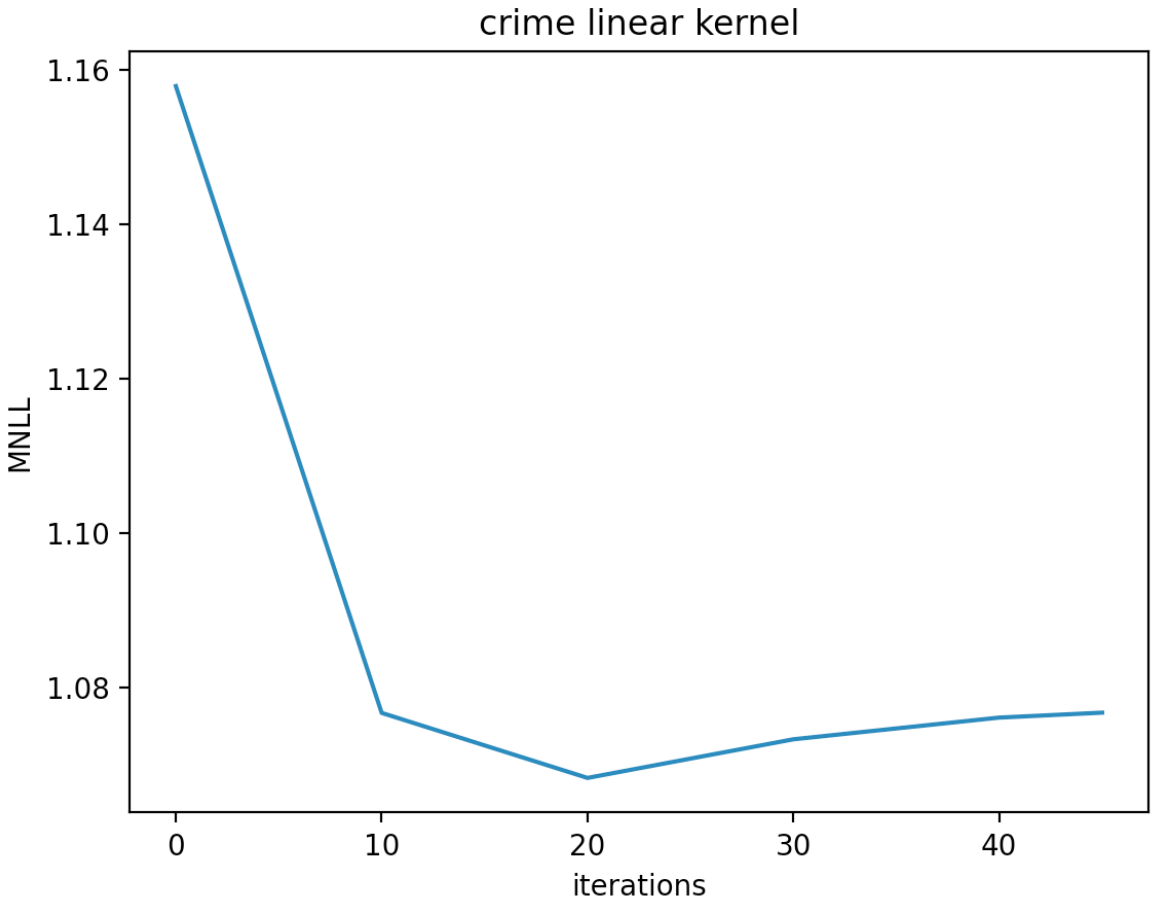
PP4 Report
Brendan McShane

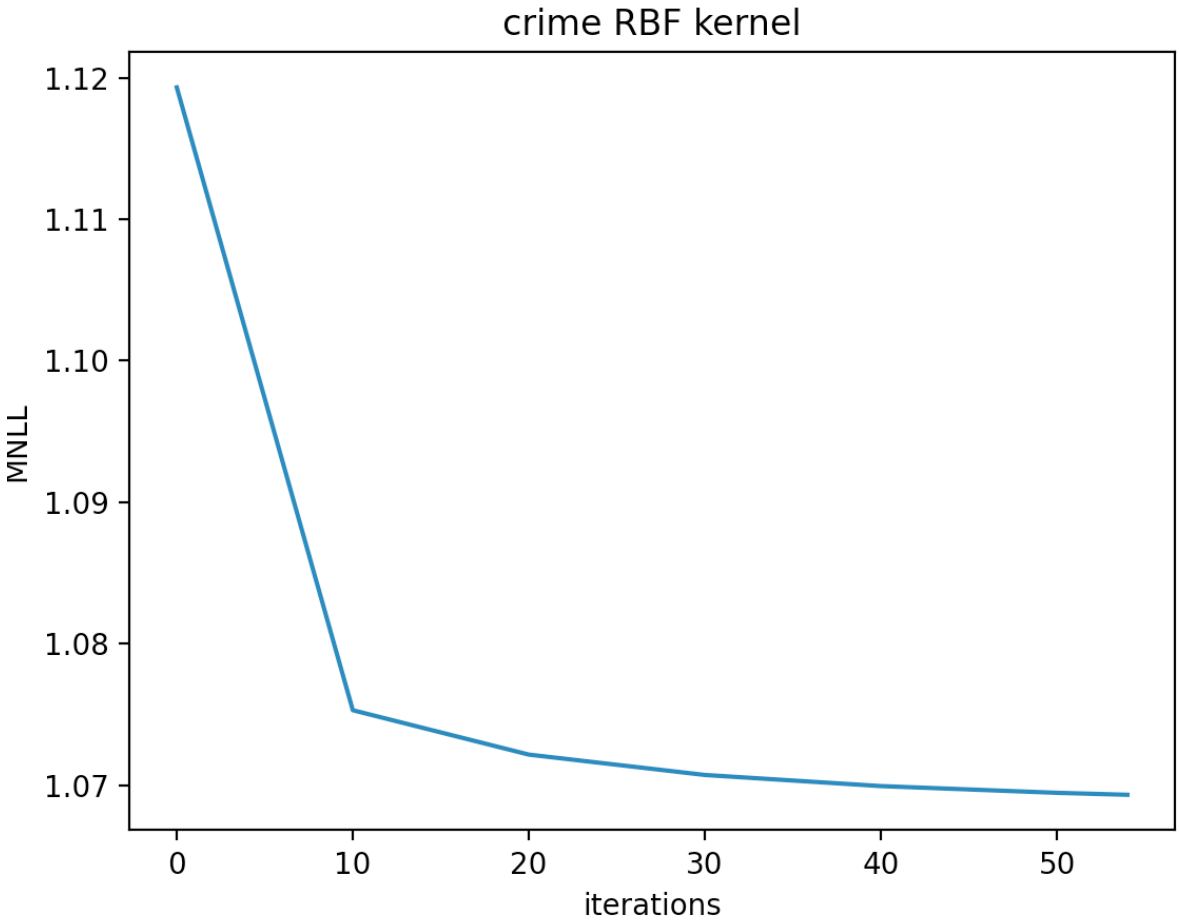
Visualizing performance on the 1D dataset

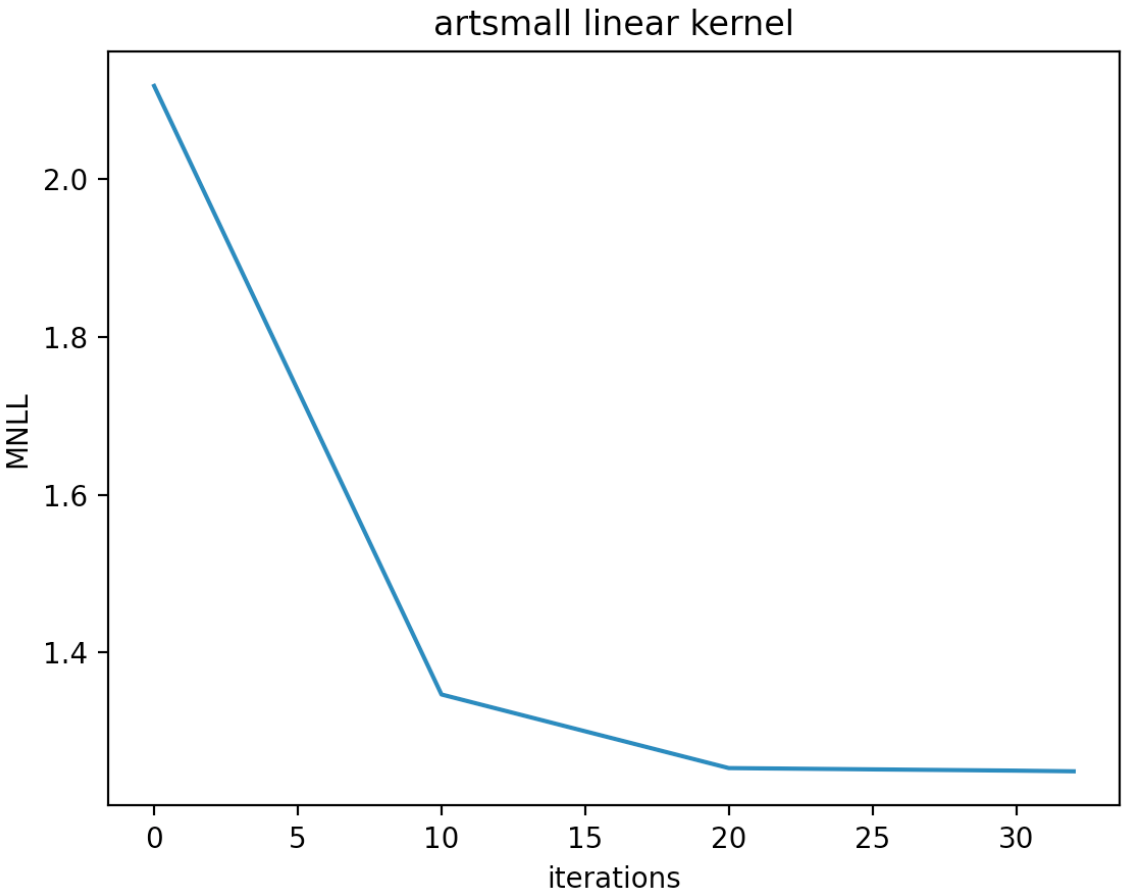


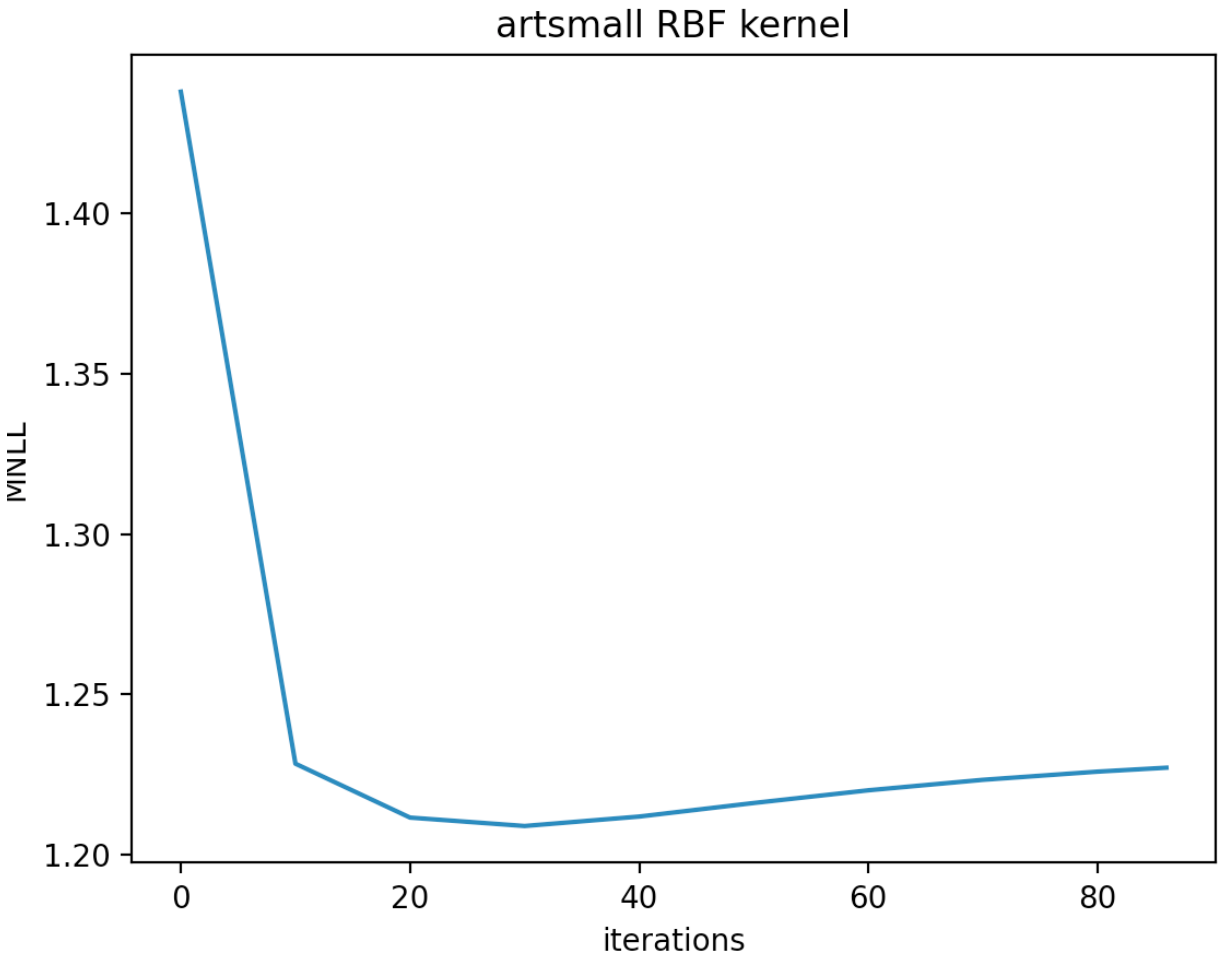


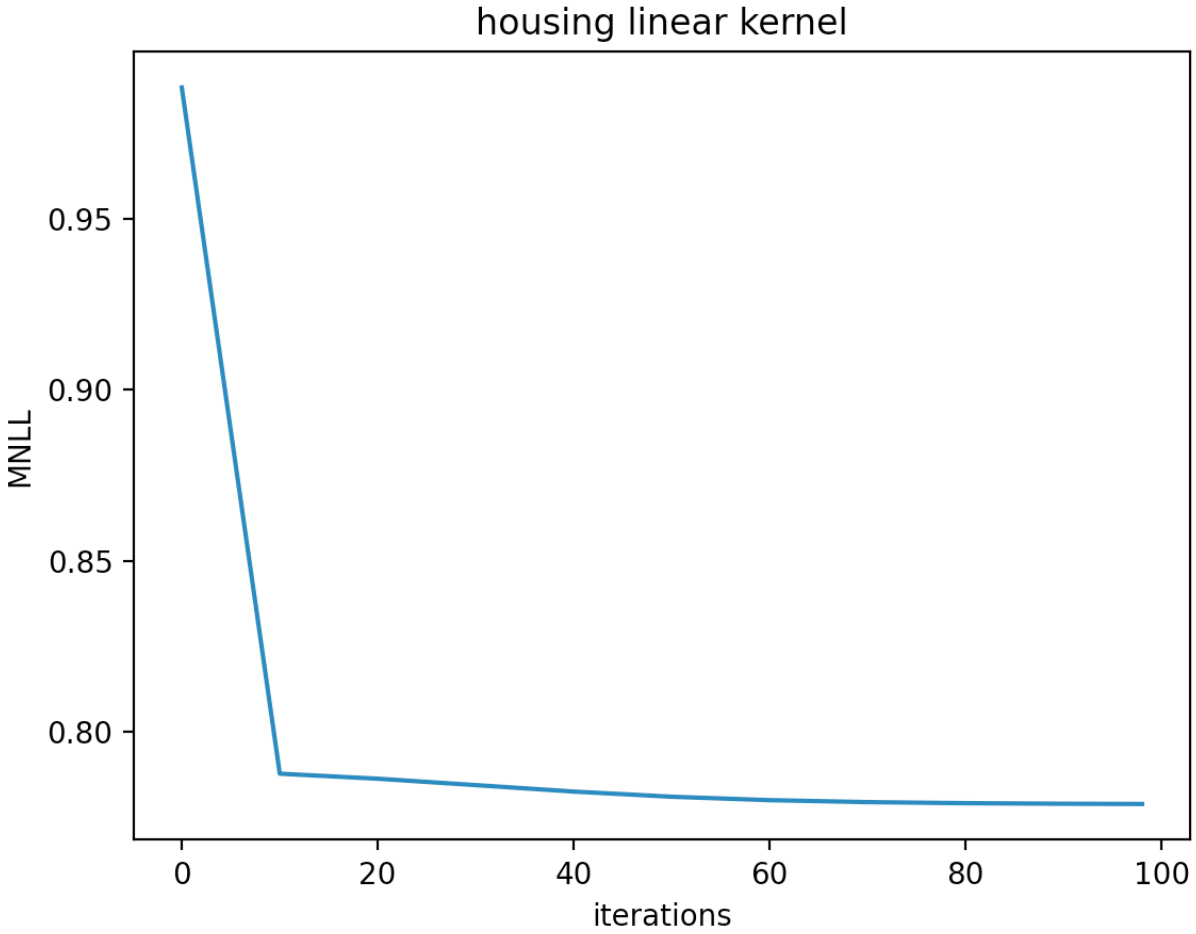
Performance as a function of iterations

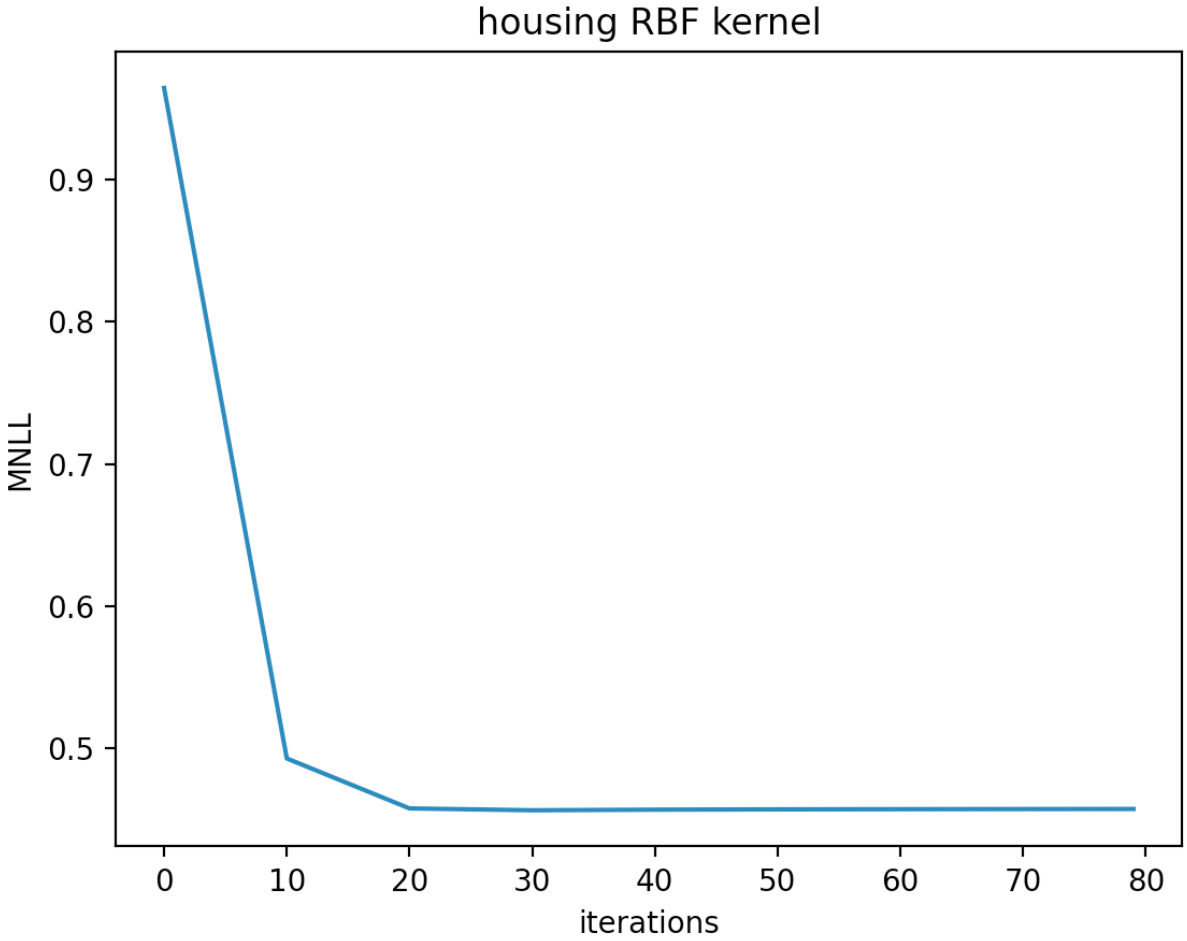

crime linear kernel
1.16
1.14
1.12
1.10
1.08
MNLL
0
10
20
30
40
iterations

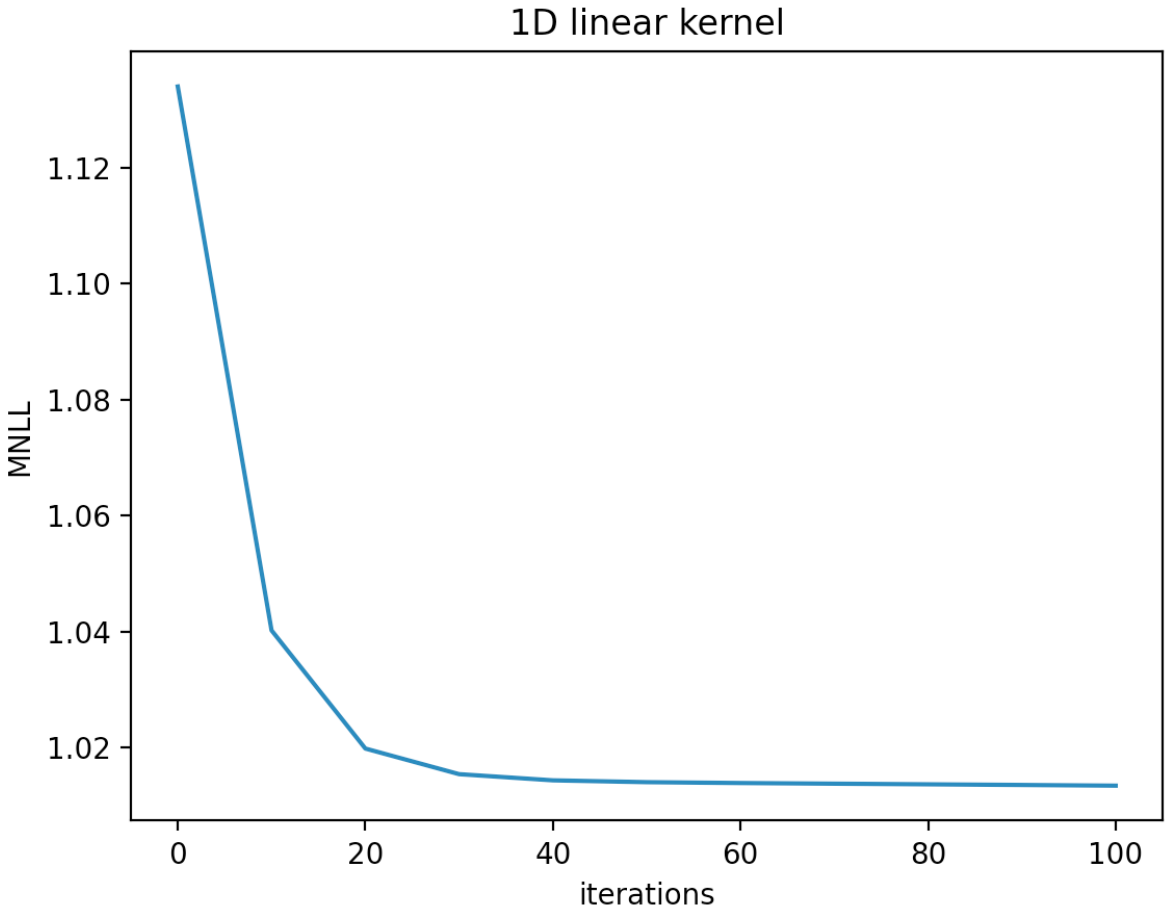

crime RBF kernel
1.12
1.11
1.10
1.09
1.08
1.07
MNLL
0
10
20
30
40
50
iterations

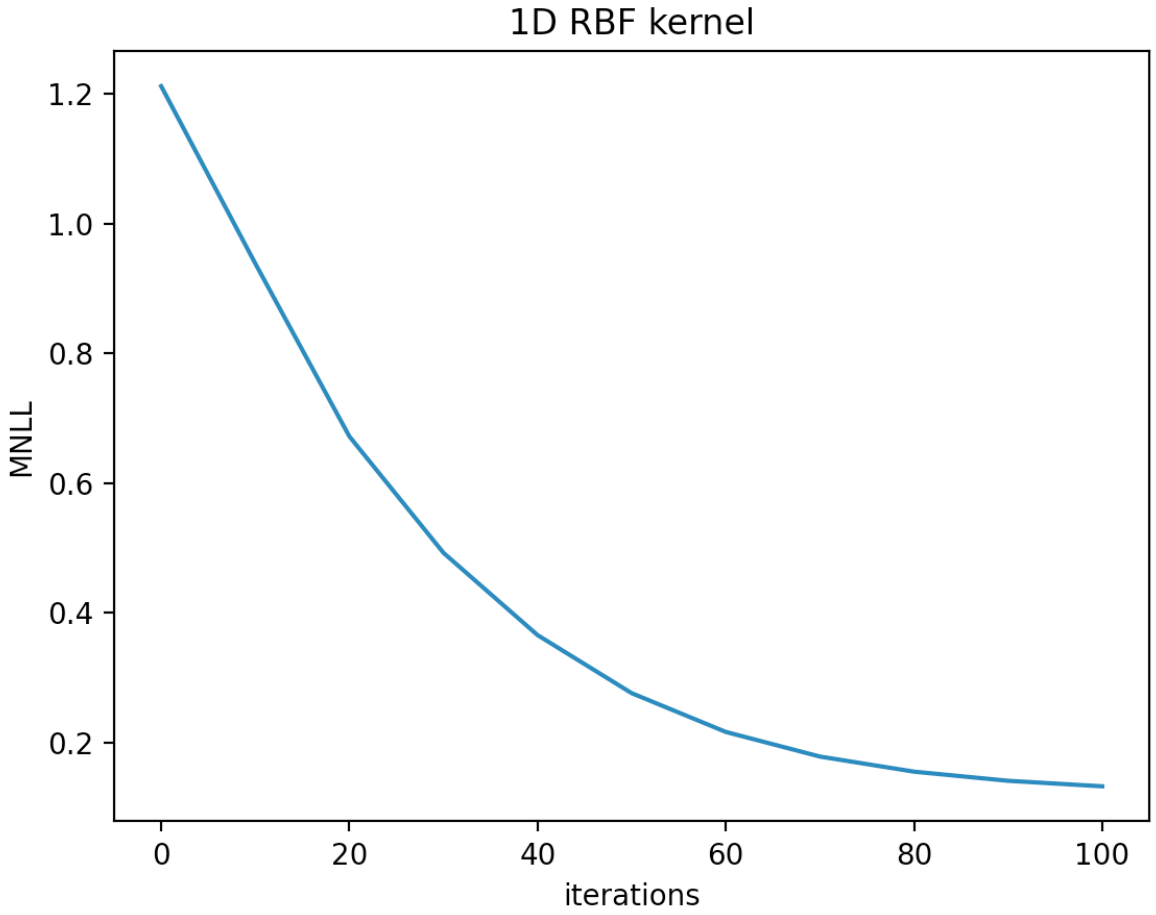

artsmall linear kernel
MNLL
2.0
1.8
1.6
1.4
0
5
10
15
20
25
30
iterations


artsmall RBF kernel
MNLL
1.40
1.35
1.30
1.25
1.20
0
20
40
60
80
iterations


housing linear kernel
MNLL
0.95
0.90
0.85
0.80
0
20
40
60
80
100
iterations


housing RBF kernel
MNLL
0.9
0.8
0.7
0.6
0.5
0
10
20
30
40
50
60
70
80
iterations


1D linear kernel
MNLL
1.12
1.10
1.08
1.06
1.04
1.02
0
20
40
60
80
100
iterations


1D RBF kernel
MNLL
iterations
1.2
1.0
0.8
0.6
0.4
0.2
0
20
40
60
80
100

Comparison to Bayesian Linear Regression

```
================== mse results
crime linear kernel
{'mse': 2.013653601729685, 'alpha': 354.79524745731277, 'beta': 2.599277226726662}

crime RBF kernel
{'mse': 2.016782317302533, 'alpha': 0.6633009955095703, 'beta': 2.756751407788210 3}

artsmall linear kernel
{'mse': 1.2691267239102149, 'alpha': 163.2361437668833, 'beta': 4.018219997282762}

artsmall RBF kernel
{'mse': 1.2134864579448654, 'alpha': 0.5183022449134349, 'beta': 6.576186014519048}

housing linear kernel
{'mse': 1.705500197578262, 'alpha': 20.871200878946215, 'beta': 3.9651437119045654}

housing RBF kernel
{'mse': 1.827199440480926, 'alpha': 0.3178114692714033, 'beta': 12.790180381542063}

1D linear kernel
{'mse': 1.1321363591621192, 'alpha': 2.4021825758505626, 'beta': 1.8619060177981634}

1D RBF kernel
{'mse': 1.5257570263168765, 'alpha': 1.1360584354500922, 'beta': 16.969963793180373}
```

dataset	MSE	α	β
crime	0.5	357.5	2.6
artsmall	0.716	141.4	4.23
housing	0.288	20.4	4.0
1D	0.39	7.5	1.9

Discussion of results

With the results from the 1D dataset we can clearly see that the square exponential kernel (aka the RBF kernel) does a much better job of learning and approximating the sin function. The RBF kernel is clearly much more flexible with what functions it's capable of learning. Most of the time, both algorithms have a sharp increase in likelihood that eventually plateaus. When the linear kernel is ran on the crime dataset or the RBF kernel is ran on artsmall, the likelihood actually eventually decreases again after the plateau. While I'm not quite sure why this happens, it implies that our model overshoots its ideal hyperparameters and we may need more sophisticated means to detect convergence on select datasets. The linear kernel appears to converge in fewer iterations on average, probably due to it having to tune only 2/3s of the parameters that RBF does, while the exponential kernel attains much higher likelihoods and more accurate models. WRT to BLR, GP seems to perform noticeably worse when it comes to MSE. Although the alpha and beta hyperparameters converge to very similar values, The MSE is much lower in BLR than it is in GP.