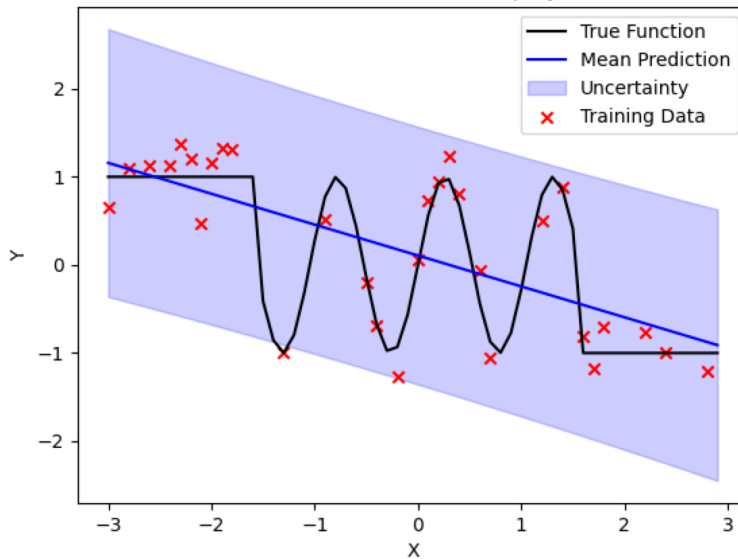Sunilsakthivel Sakthi Velavan
Roni Khardon
CSCI-B 555
20 Nov 2023
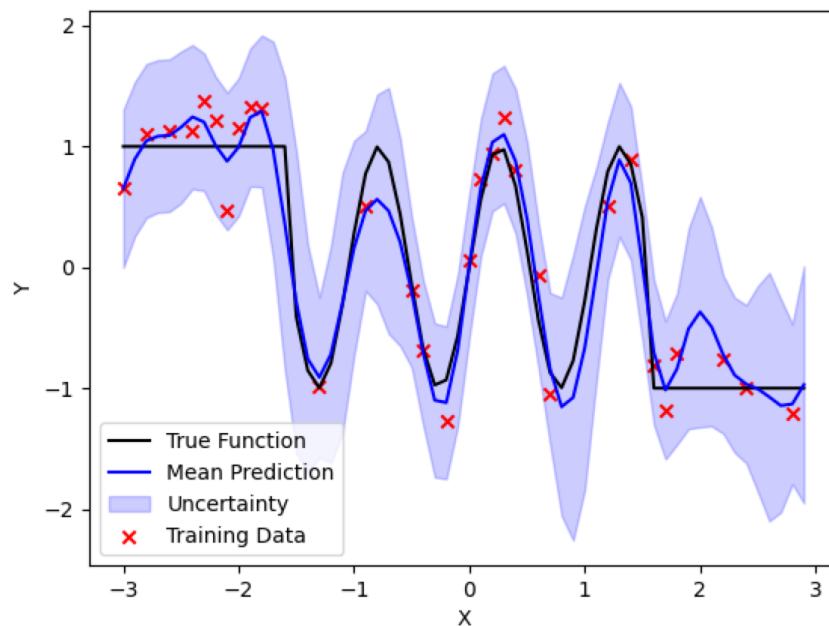
Programming Project 4

Visualizing performance on the 1D dataset:

GP Predictive Distribution - 1D Dataset with polynomial kernel function
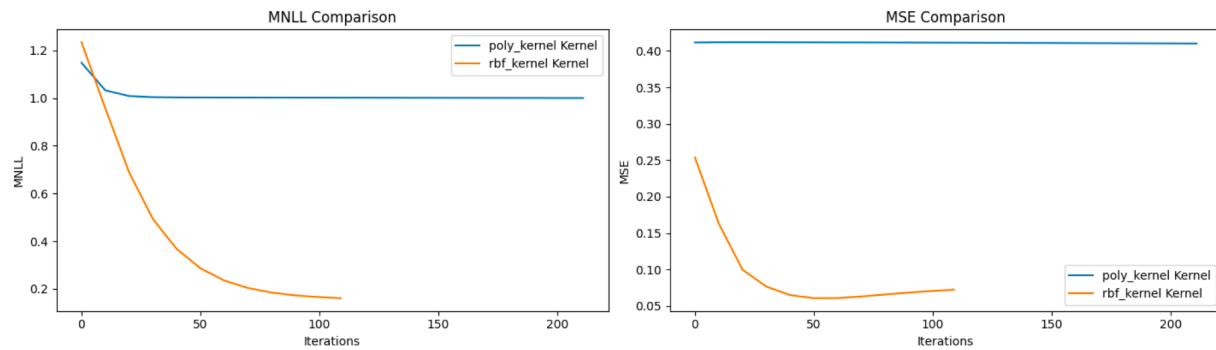

GP Predictive Distribution - 1D Dataset with rbf kernel function

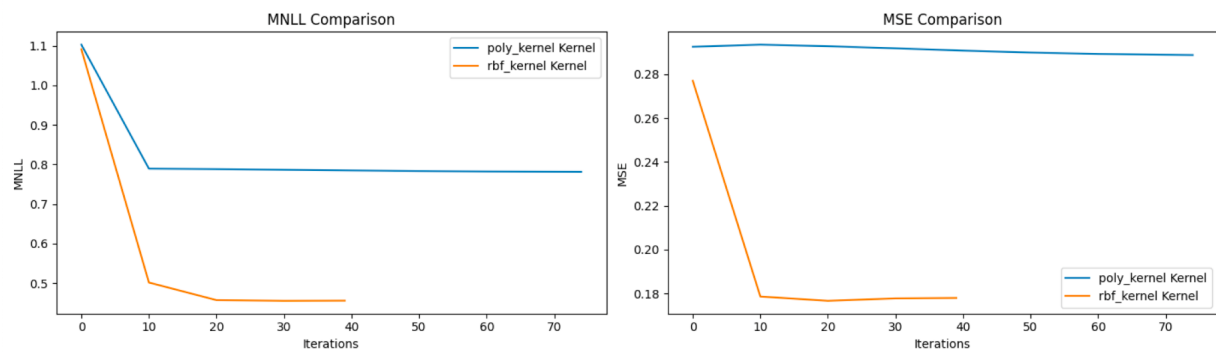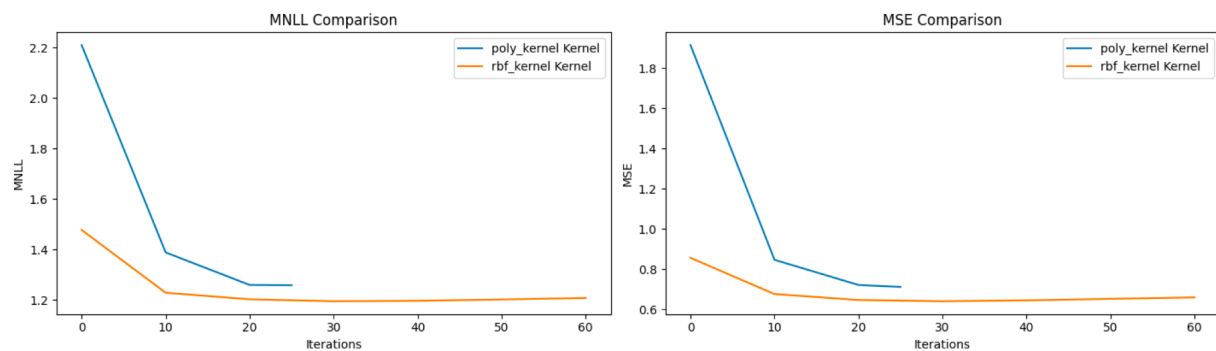# Performance as a function of Iterations:
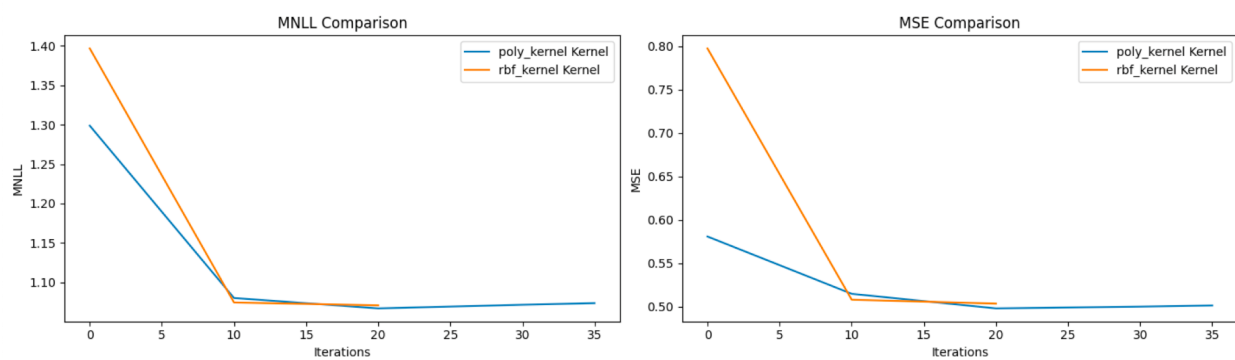
Performance Comparison between Kernels - 1D

Performance Comparison between Kernels - housing

Performance Comparison between Kernels - artsmall

Performance Comparison between Kernels - crime

Comparison of Bayesian Linear Regression:

Gaussian Process Results:

Dataset	MSE	alpha	beta
crime	0.50151866890626	320.669764901518	2.61138271021256
artsmall	0.71240591135867	140.89509999224	4.13443167341271
housing	0.28866544127906	18.7018324256571	4.00469585550641
1D	0.41025834394749	5.25808956496352	1.92914059346791

Bayesian Linear Regression Results:

Dataset	MSE	alpha	beta
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:

The recorded Mean Squared Error (MSE) values for both Bayesian Linear Regression (BLR) and Gaussian Processes (GP) with a linear kernel demonstrate a noteworthy similarity. In general, the two algorithms exhibit comparable test errors across different datasets. This consistency in performance indicates that, despite differences in their underlying principles and the way we generate the hyperparameters, both models are adept at capturing the relationships within the given datasets. As one can tell the MSE values on Gaussian Process pretty closely reflect the MSE values in the BLR method. We can also see that, for the most part, both the alpha and beta values in Gaussian Process seem to mirror the BLR values. We do notice that for the crime and 1D dataset, the alpha values seem to deviate from the BLR alphas. While this isn't necessarily a bad thing considering that the MSE is more or less the same, The deviation in alpha values for the crime and 1D datasets suggests a nuanced relationship between the Gaussian Process hyperparameter tuning and the underlying characteristics of the data. This divergence may stem from the inherent differences in the modeling assumptions and methodologies between the two approaches.

As for the RBF kernel, we can take a closer look at the graphs generated as part of the "Performance as a function of iterations" to see the recorded Mean Negative Log Likelihood (MNLL) and the MSE values given the selected alpha, beta, and s and the respective predicted values they generated. A notable trend emerges where the RBF kernel consistently converges at a lower error value than the polynomial (poly) kernel, except for the crime dataset. This divergence in performance prompts a closer inspection into the intrinsic qualities of the crime dataset that challenge the RBF kernel's conventional superiority. Moreover, the RBF kernel consistently achieves its lowest error at fewer iterations compared to the linear kernel, except for the artsmall dataset, where the flexibility of the RBF kernel seems to require additional iterations for convergence. This observation underscores the importance of tailoring kernel choices to specific dataset characteristics. The RBF kernel, with its capacity to capture non-linear relationships, excels in datasets demanding a more complex model, while the linear kernel remains proficient in situations where linearity is predominant.

While the linear kernel offers computational efficiency and interpretability, it may falter in the face of intricate relationships present in datasets such as artsmall. On the other hand, the RBF kernel's enhanced flexibility comes at the cost of increased computational complexity, making it crucial to weigh the trade-offs based on the specific demands of the given dataset.