3.)

a.)

See photo RegularizedLinearRegression.

degree1

avgLoss1.29165089

degree2

avgLoss0.93446707

degree3

avgLoss0.06539581

degree4

avgLoss0.06024950

Best results were achieved at degree 4.

When the degree of the monomial basis function is d, the dimension of the feature space is (d^2 + 3d + 2)/2. We invert a matrix of this size, and thus the runtime is O((d^2)^3) = O(d^6).

b.)

See photo BayesianLinearRegression.

degree1

avgLoss1.29165089

degree2

avgLoss0.93446707

degree3

avgLoss0.06539581

degree4

avgLoss0.06024950

Best results were achieved at degree 4.

When the degree of the monomial basis function is d, the dimension of the feature space is (d^2 + 3d + 2)/2. We invert a matrix of this size, and thus the runtime is O((d^2)^3) = O(d^6).

Due to the selection of parameters in this problem, generalized Bayesian linear regression is the exact same algorithm as generalized linear regression. Thus, the results and runtime analysis are the exact same.

c.)

For the identity kernel, Mean Squared Error=3.5758

For the Gaussian kernel, see photo Gaussian Kernel.

For the Polynomial kernel, see photo Polynomial Kernel.

In all choices of kernel, the most expensive operation was inverting a matrix the size of the training data set. Thus, the runtime in each variation is cubic in the size of the data set.

The Gaussian and Polynomial Kernels naturally performed better due to having a larger feature space.