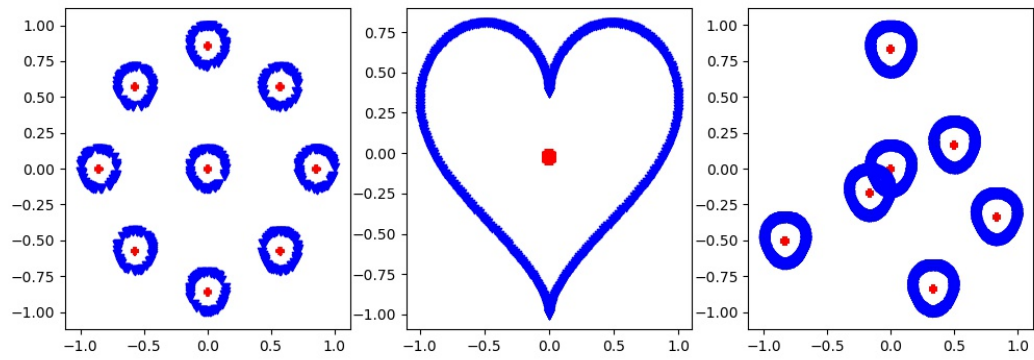
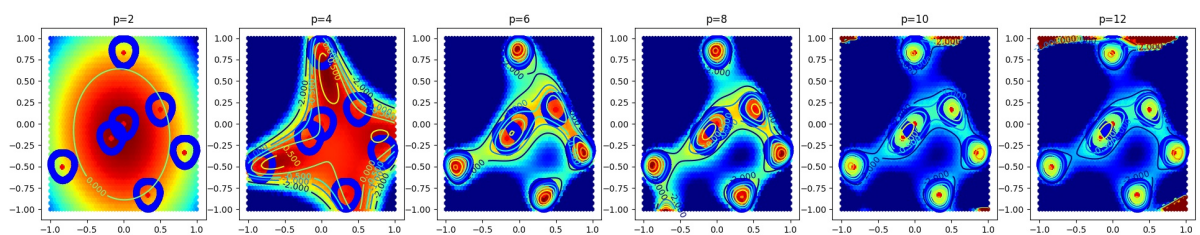
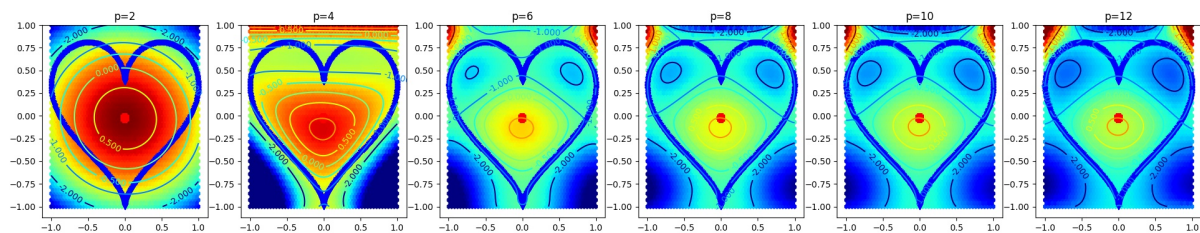
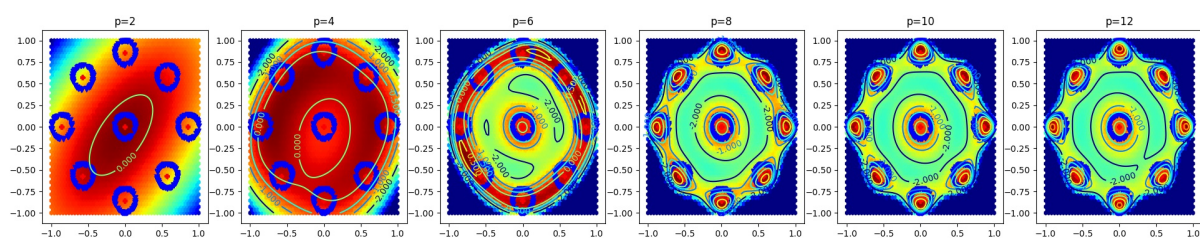
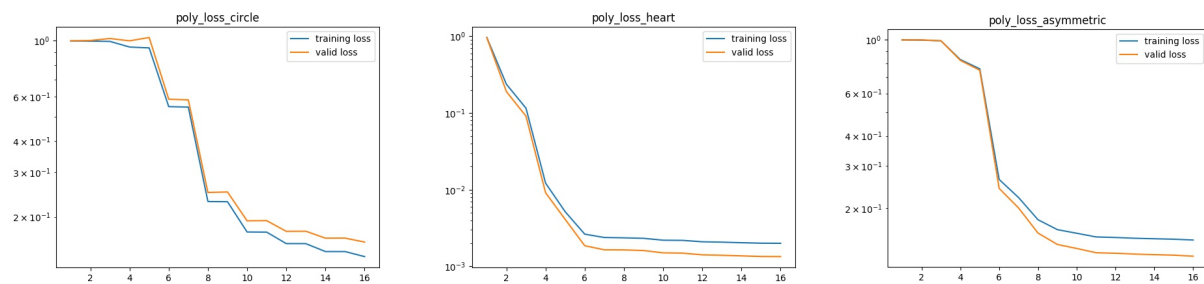


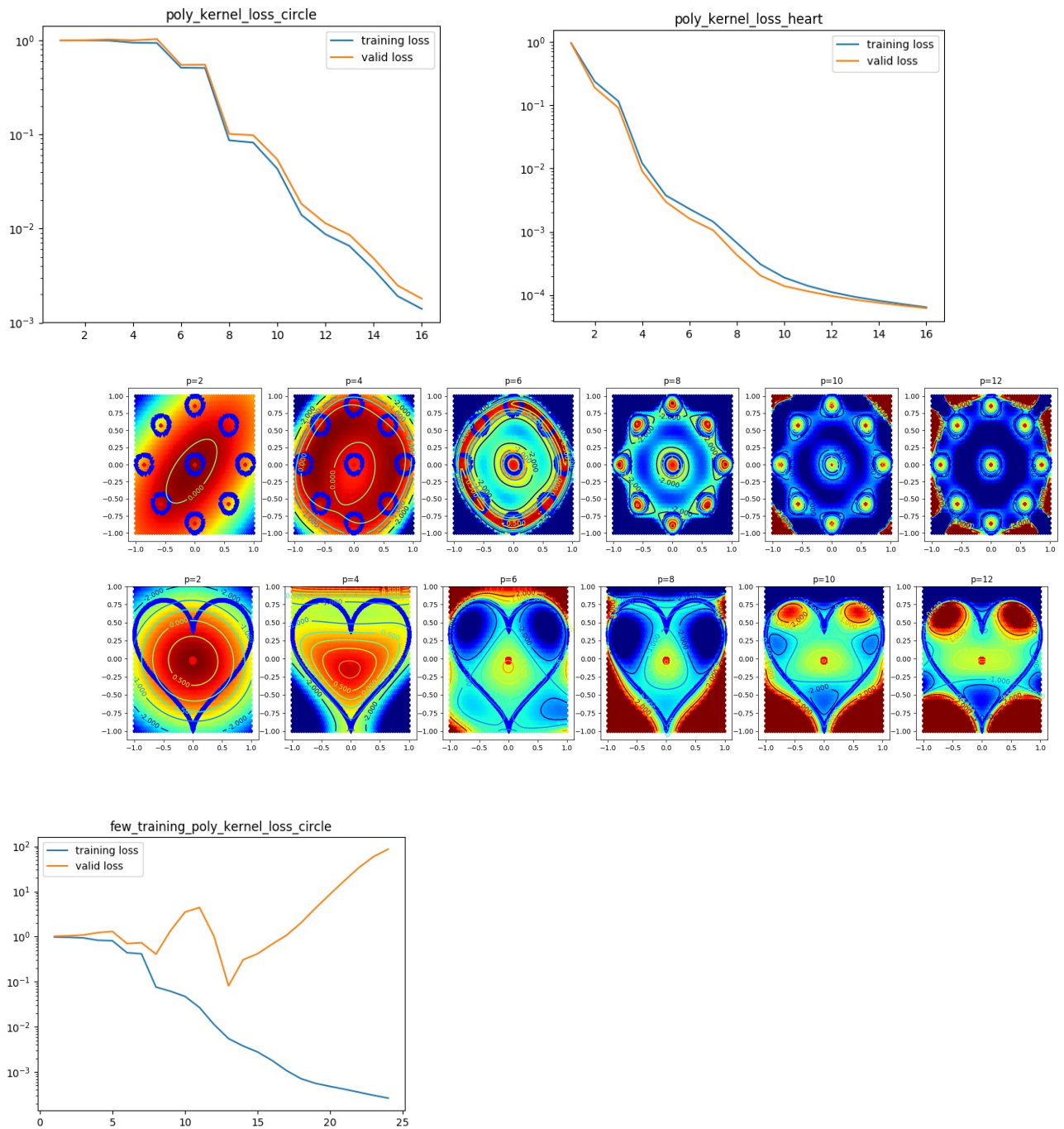
5a



5b

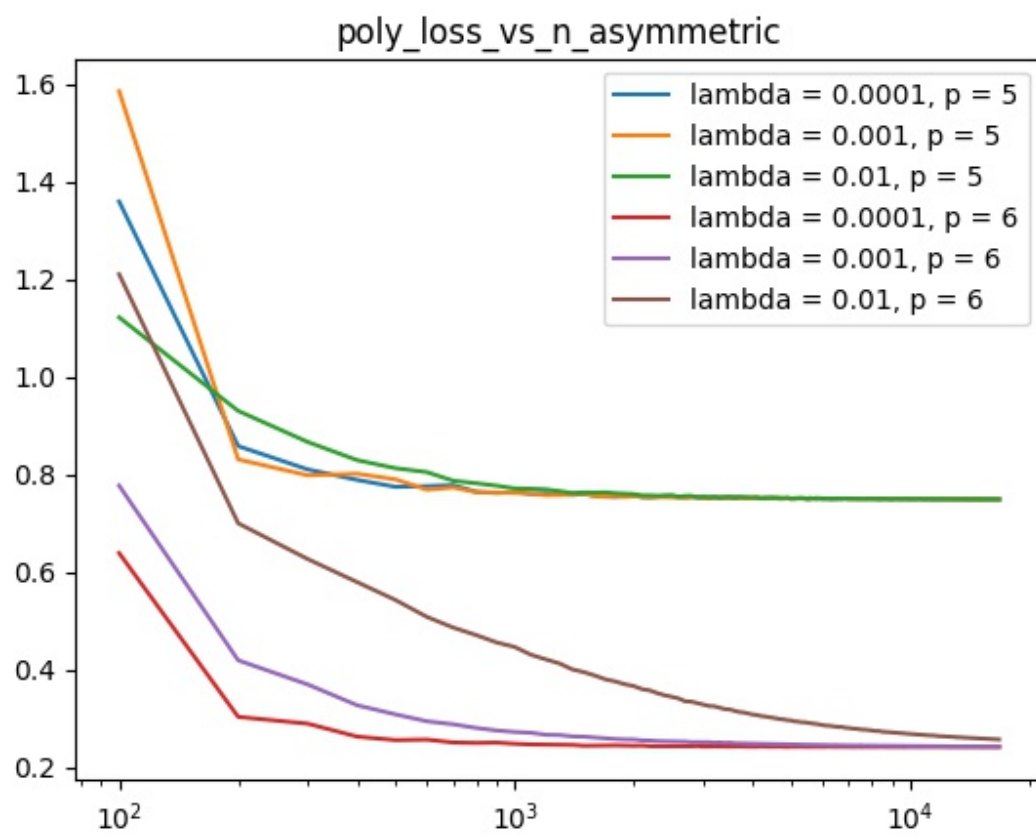


5c

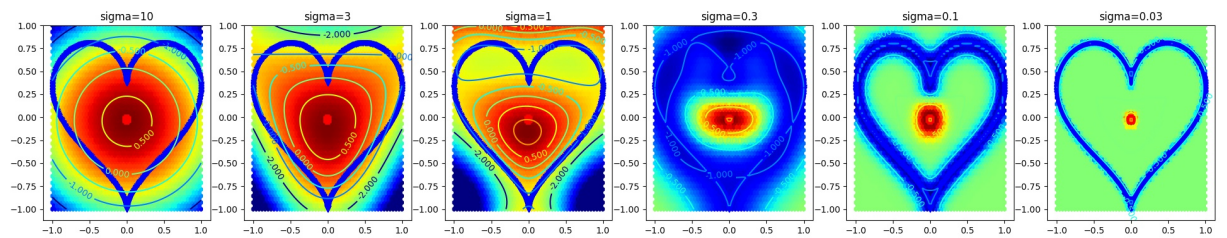
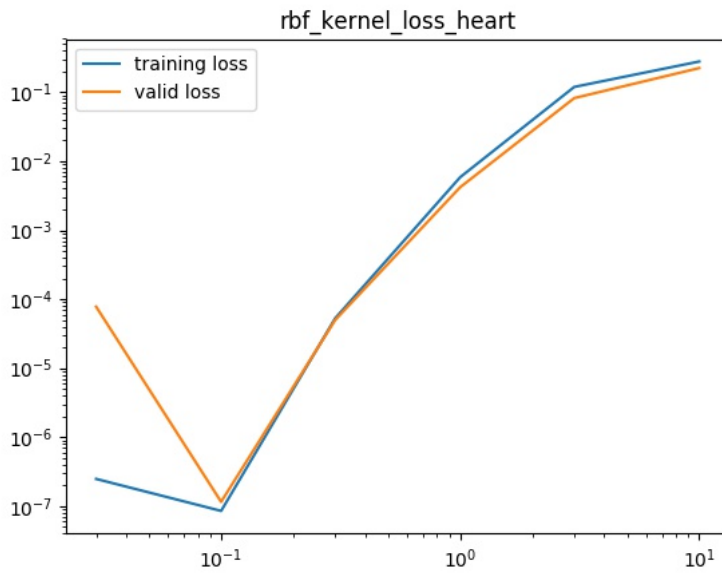


We only want high-order polynomial when we have a lot of training data.

5d



5e



validation error is the smallest at $\sigma = 0.1$. Validation error first decreases and then increases as σ increases.

5f

When $p = 16$, # of features = 153, which is much smaller than the number of points, so the non-kernelized one should be more efficient.

It is not possible to implement rbf as non-kernelized ridge regression, since if we do Taylor expansion for the exponential function, there are infinite terms, which means the number of features is infinite.

We would prefer kernel ridge when the number of polynomial features are much larger than the number of samples.

5g

For polynomial kernel, if we don't clip, we can see that as polynomial order increases, the polynomial explodes very quickly. RBF kernel doesn't have this problem, since the RBF kernel is always smaller than 1.

