COMS W4903: Machine Learning for Data Science
Homework 3

jc4609

Problem1：

a)

I set b = 1 and σ^2 = 0.1.

the means of the prediction of 42 test data are:

-3.92179075 -7.4345191 7.60996334 -6.31367723 -2.62707515
  2.42672863 -5.10105538 -1.04150101 -9.68122146 11.55274558
 -7.3462377 14.84230235 -9.14838384 -1.95397238 -0.4338344
 -7.50431681 -2.2415783 1.34697488 10.82486549 -5.109959
 -9.38444244 12.02812522 -2.95472182 -5.07718027 -8.69914078
 -8.20284863 -0.01686867 1.68863614 -1.64833507 -4.41897967
 -6.58242232 -3.86901518 2.08305269 -2.17138509 -4.08778533
  4.42184086 8.33097311 14.13907744 9.68195815 -2.91138573
  2.62845625 -1.62617197

And the variances are:

0.14312434 0.31913822 0.15028643 0.23651514 0.62608229 0.13772178
 0.29523625 0.33026309 0.28532645 0.15407895 0.140222   0.24644814
 0.15251094 0.31513015 0.71264284 0.40643616 0.31666717 0.14987994
 0.1227873 0.47017533 0.18972224 0.12615329 0.17832243 0.64389399
 0.36276091 0.16873554 0.54762229 0.14410095 0.19144561 0.18551819
 0.66571585 0.85389603 0.18411343 0.6873196 0.15895114 0.20924173
 0.17775119 0.12816991 0.12994271 0.31425854 0.16495329 0.36569658

b)

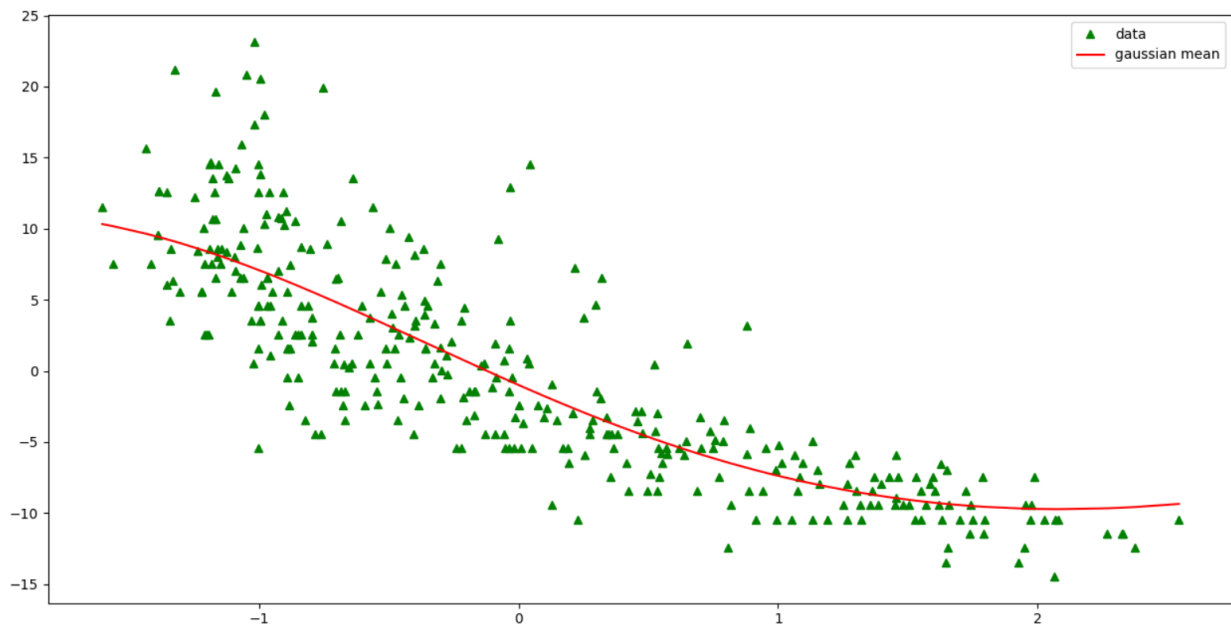| σ^2 / b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
5	1.966	1.933	1.923	1.922	1.924	1.929	1.934	1.940	1.946	1.953
7	1.920	1.905	1.908	1.916	1.924	1.933	1.942	1.950	1.958	1.965
9	1.898	1.903	1.918	1.933	1.946	1.957	1.96	1.976	1.985	1.992
11	1.891	1.919	1.938	1.957	1.973	1.986	1.996	2.006	2.014	2.021
13	1.896	1.936	1.965	1.986	2.001	2.014	2.024	2.033	2.041	2.049
15	1.910	1.960	1.990	2.012	2.027	2.039	2.049	2.058	2.066	2.073

c)

I get the smallest RMSE when b = 11 and σ^2=0.1, and RMSE = 1.891. The smallest RMSE of first homework's algorithm is about 4.5. Thus the Gaussian process is much better.

The first drawback of Gaussian process is that the computation volume of Gaussian process is greater than that of ridge regression, because instead of calculating xi*xj , we have to calculate exp(||xi-xj||^2) which is more complex.
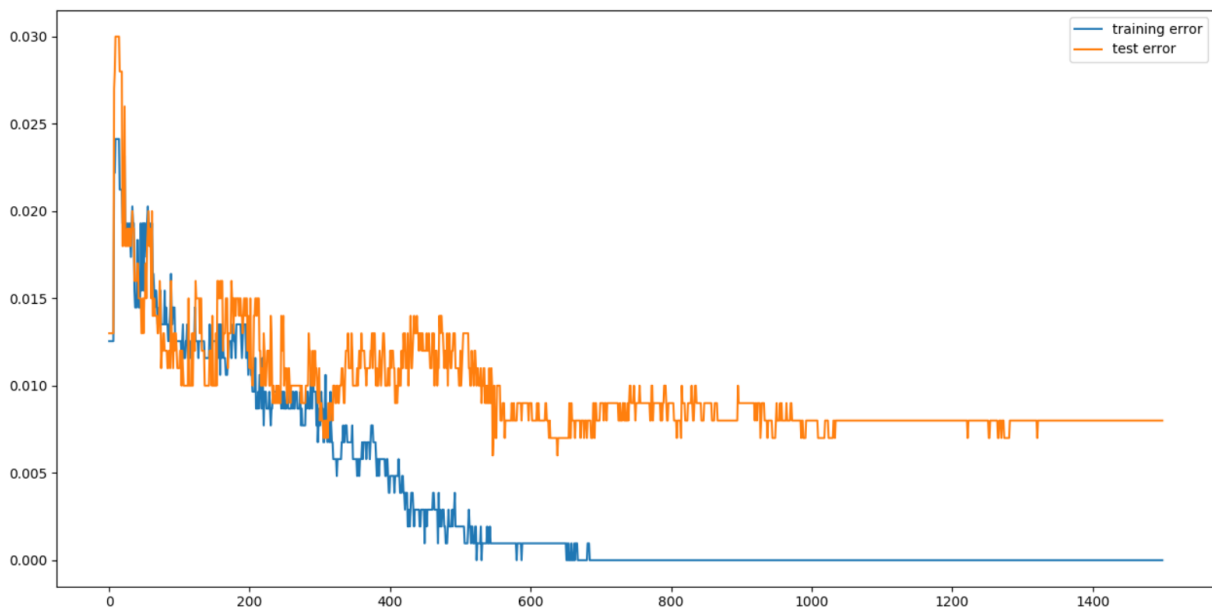
The another possible drawback is that Gaussian process is more prone to overmatch than ridge regression. The reason is that Gaussian process map the data into infinite high dimension space so it has higher possibility to overmatch the data.
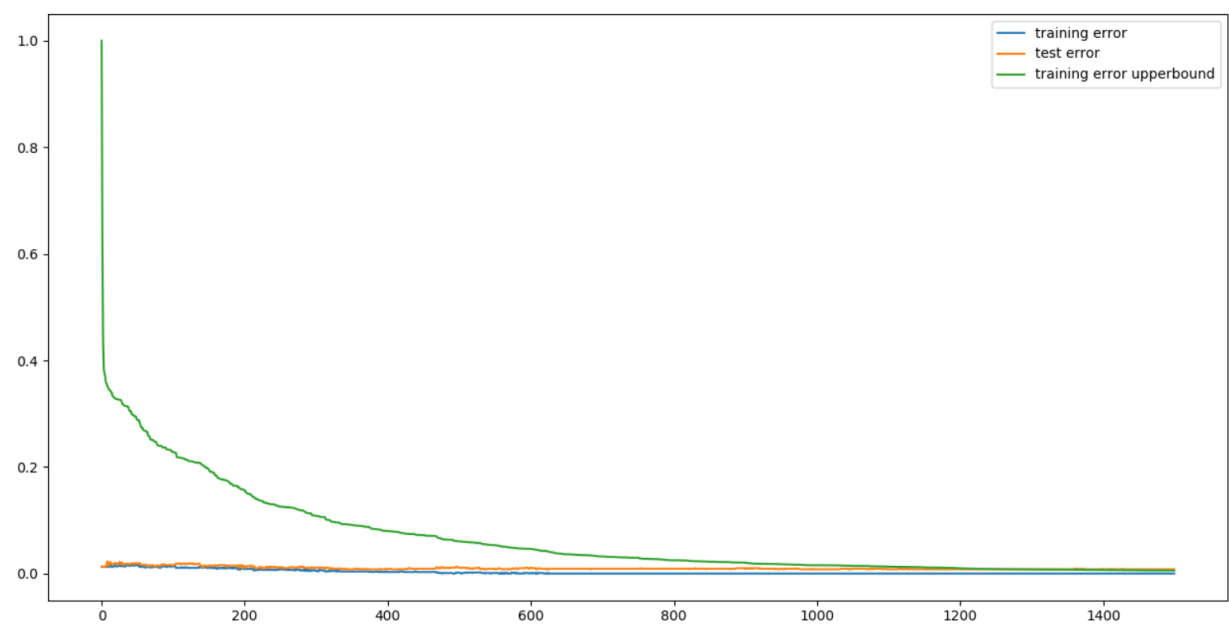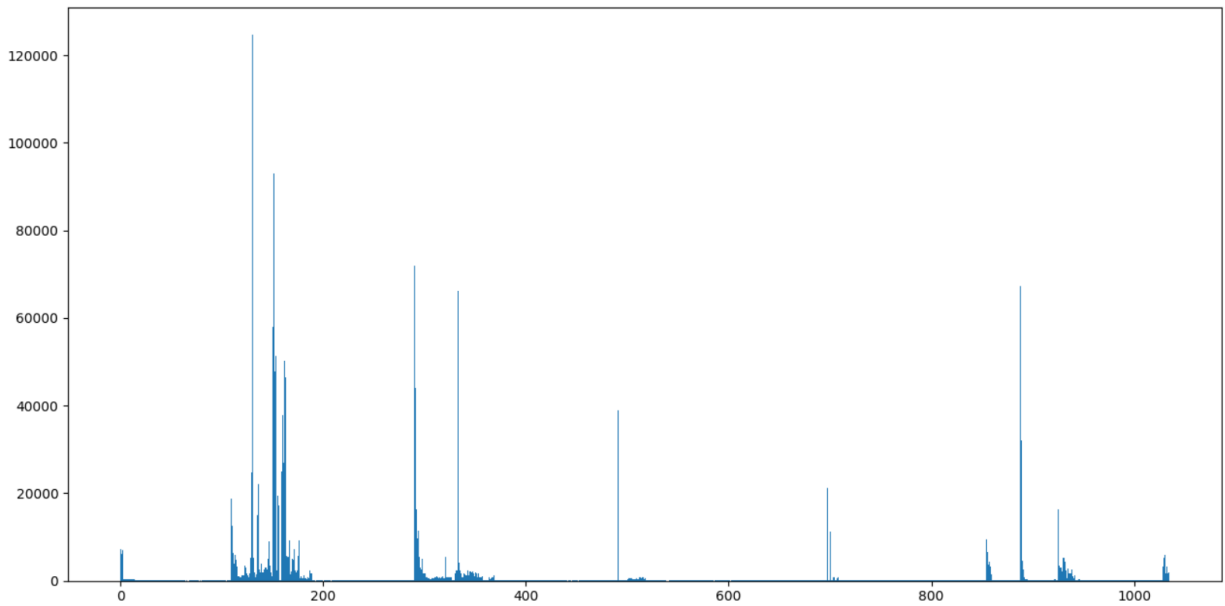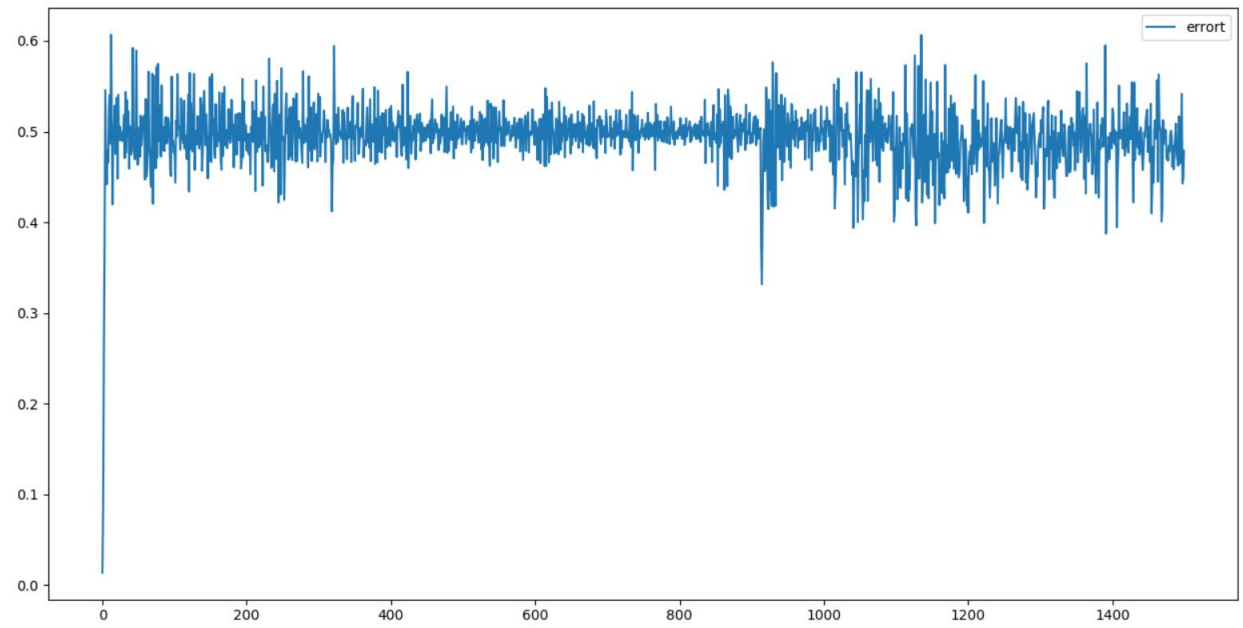
d)

Problem2：
a)

b)

c)

d)
ϵ_t:

α_t: