Marthey Antoine (am4665)
Machine Learning Homework 3

EXERCICE 1 :

- Question 1

See code

- Question 2

RMSE Values

	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
5	1.966276	1.93313534	1.9234202	1.9221976	1.92476912	1.92921255	1.93463407	1.9405831	1.94681994	1.9532125
7	1.920163 26	1.904876 52	1.90808033	1.91590168	1.924804 08	1.9337014	1.942253 77	1.950380 04	1.95809307	1.965438 04
9	1.897648 75	1.902519 13	1.91764769	1.93251433	1.945699 47	1.95723479	1.967403 26	1.976491 75	1.98474077	1.992341 21
11	1.890507 11	1.914981 01	1.93884878	1.95793625	1.973215 82	1.98576421	1.996375 14	2.005603 22	2.01383544	2.021344 8
13	1.895848 5	1.935585 9	1.96459732	1.98550199	2.001314 27	2.01387845	2.024310 38	2.033306 78	2.04131749	2.048641 55
15	1.909603 22	1.959548 83	1.99080359	2.01191545	2.027370 28	2.03946517	2.049463 38	2.058104 89	2.06584527	2.072976 06

- Question 3

The min of the table is 1.89050711498, data for b = 11 and sigma = 0.1

Unfortunately, I can't compare it to the first Homework since I didn't get the right results. My assumption is that the Gaussian Processes return a better result.

Drawback of the method: when we will have a huge dataset then the algorithm will run in a very long time, since the Kernel needs to compare each point to each other. So the method is great for a small dataset and not appropriate for a huge dataset.

- Question 4

See figure 1

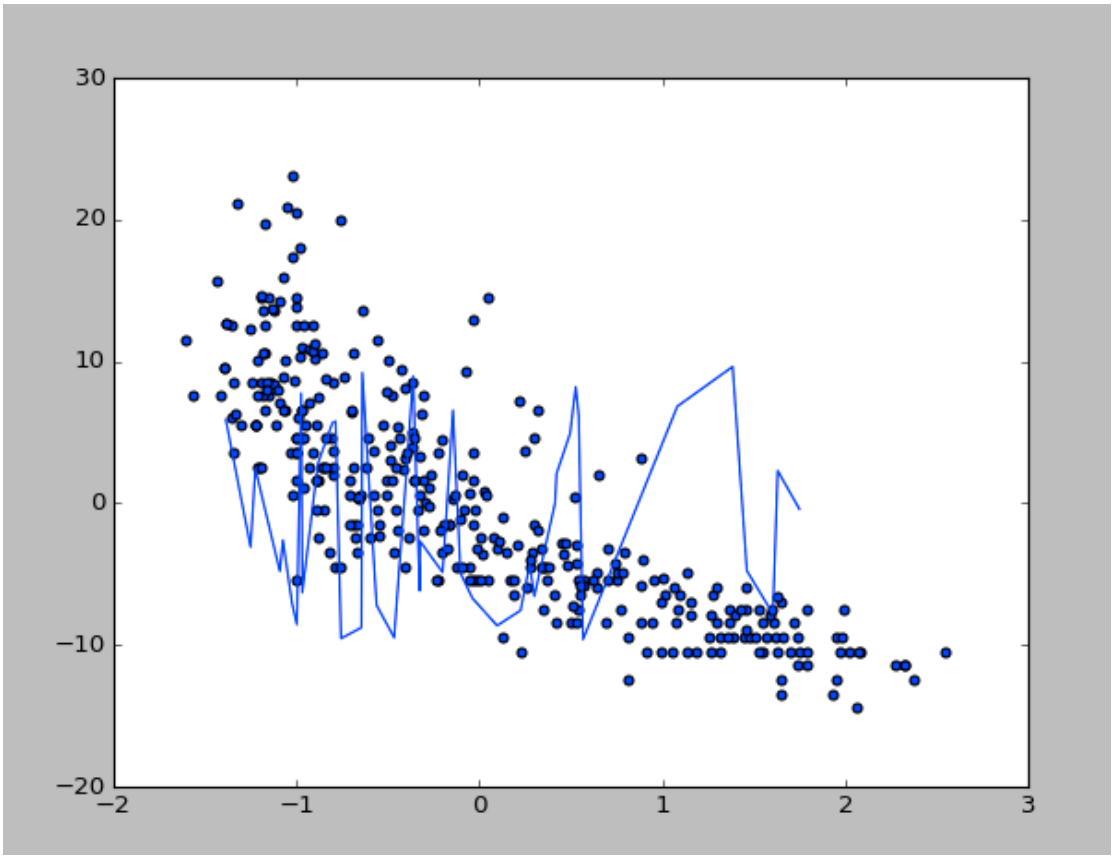

Figure 1

EXERCICE 2 :

- Question 1

See figure 2

The Green line is the testing error.
The Blue line is the training error.

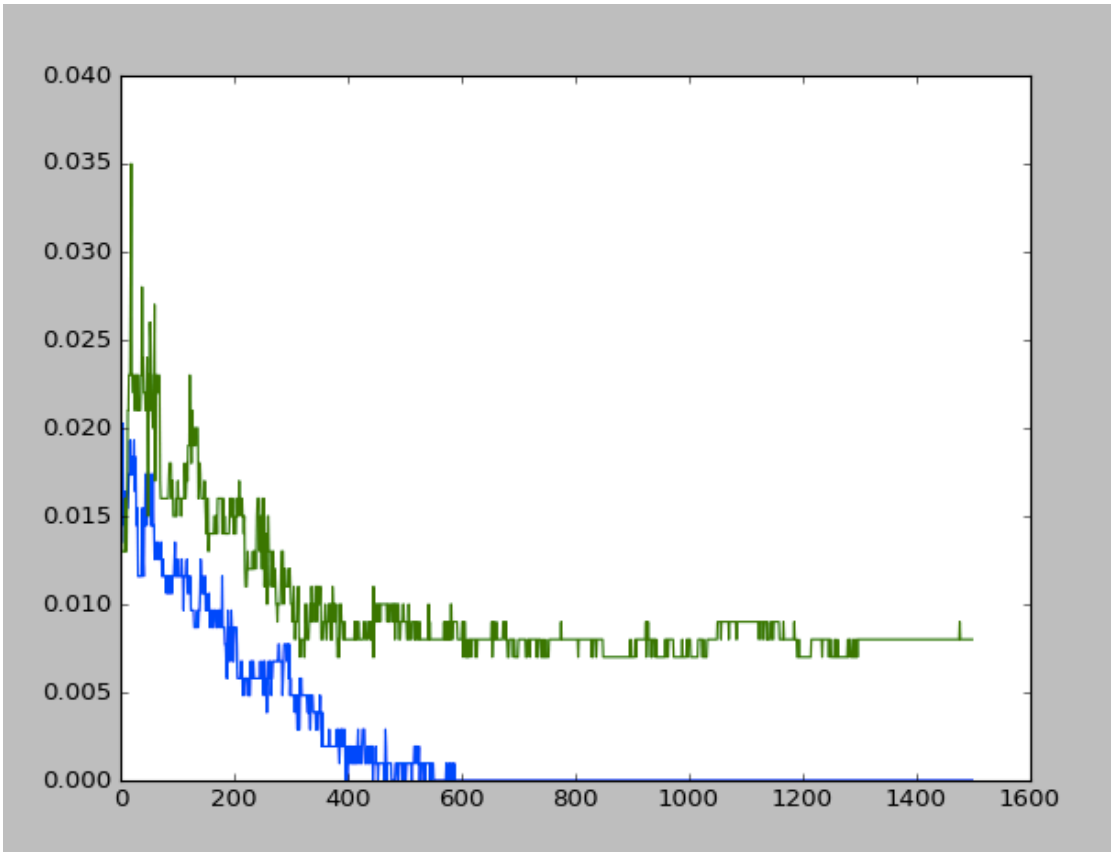

Figure 2

- Question 2 :

See Figure 3.
The Blue line is the upper bond of the training error

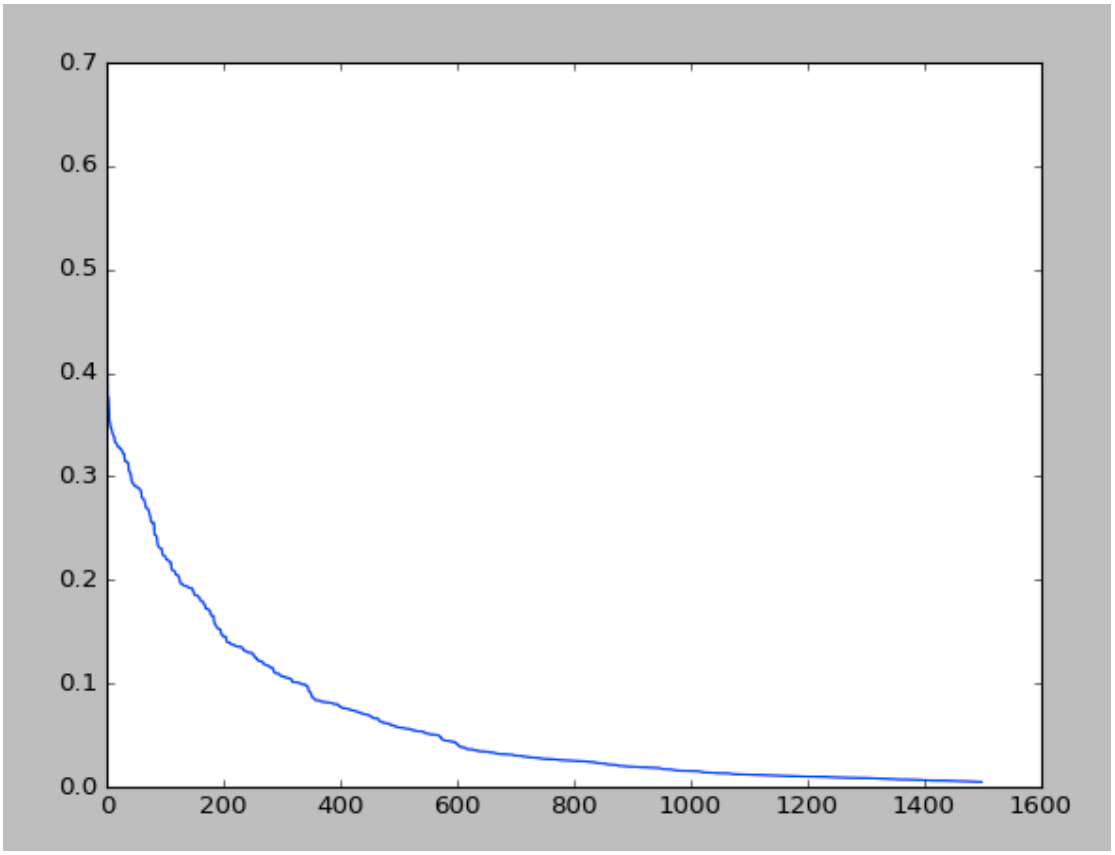

Figure 3

- Question 3

See Figure 4.
It represents the number of occurrence of the points in the differents t bootstraps.

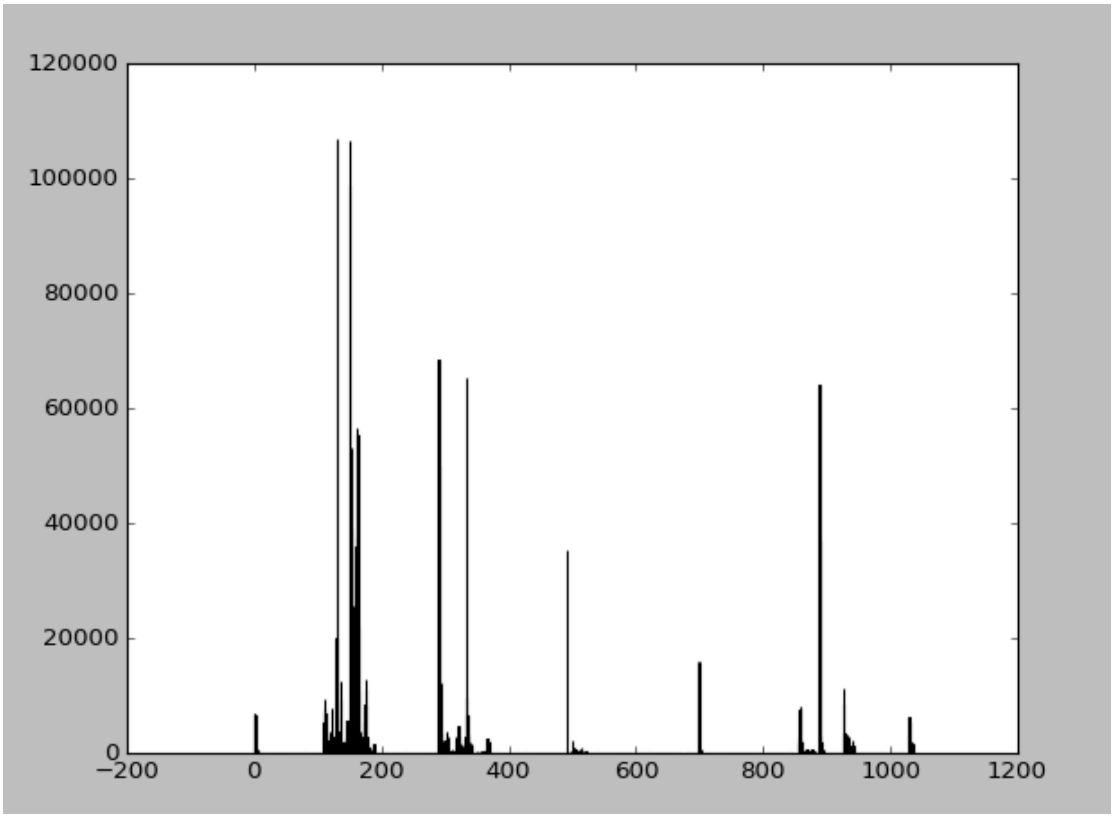

Figure 4

- Question 4 :

See Figure 5. Figure 5 represents alpha as a function of t.
See Figure 6. Figure 6 represents the error as a function of t.

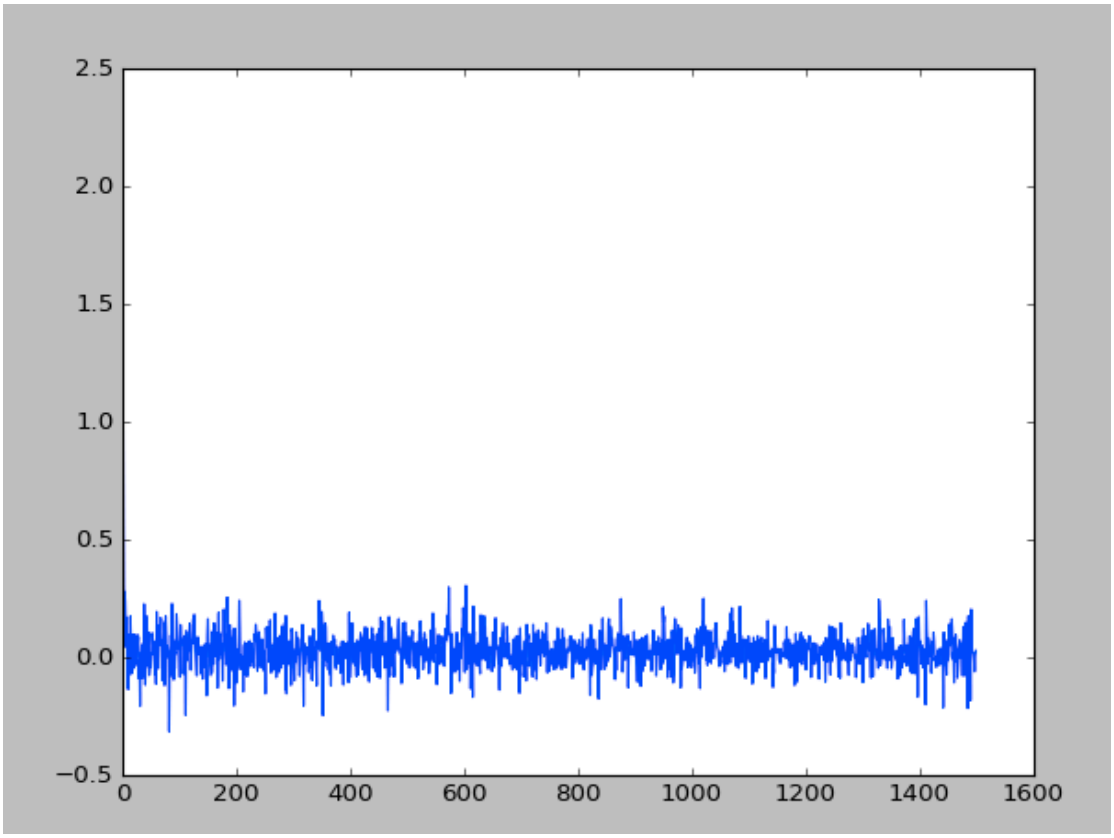

Figure 5

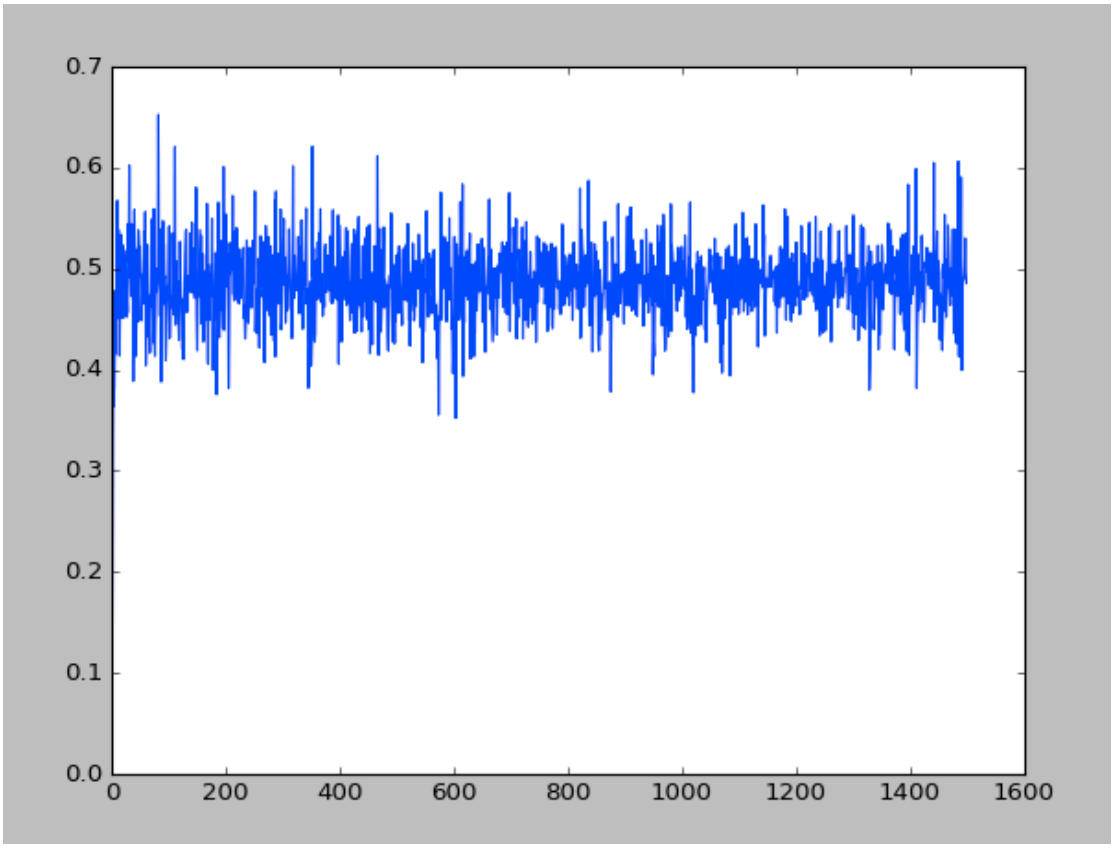

Figure 6