Trang Nguyen
Si Thu Aung

I. Comparison and Statistical Test:

Running cross validation to compute mean squared error for both Linear Regression method and Gaussian Process method with exactly same splits (how we split the data into test and training set). We get the following result

	Mean Square Error	
Split no.	GP with SE kernel (A)	Linear Regression (B)
1	2.3433e-8	2.3610e-8
2	7.9063e-8	6.5224e-8
3	5.9323e-5	1.1699e-7
4	2.3460e-6	5.0202e-8
5	1.2397e-4	3.1524e-8
6	1.0365e-6	1.0081e-7
7	1.7923e-6	1.1891e-7
8	1.9382e-5	1.4481e-7
9	7.3891e-7	1.0898e-7
10	6.2197e-6	1.8753e-7

1. t -test:

In order to determine which model is better, we can do a parametric test. We have that

$$\bar{d} = \mu_A - \mu_B$$

$$\sigma_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$$

Hence, we have $\bar{d} = 2.14 * 10^{-5}$ and $\sigma_d = 4.04 * 10^{-5}$. Having that, we can find t by using the equation

$$t = \frac{\bar{d} - \mu_0}{\sigma_d/\sqrt{n}},$$

With $\mu_0 = 0$, we have $t =$ **1.67**

We know that critical value of t distribution with two tails at $\alpha = 0.05$ and $df = 9$ is **2.262**.

We have that $t < t_\alpha$, so we cannot reject H_0 at significance level α. Therefore, the parametric test tells us that the two methods performs with the same accuracy.

2. Sign test

Number of splits that GP performs better is $A_{win} = 1$
Number of splits that LG performs better is $B_{win} = 9$
We use two-tails version of the test with the level of significant to be $\alpha = 0.05.$ We found that the number of wins have to be larger than 8 to be consider significant better. Hence, in our model, the sign test tells us that Linear Regression performs significantly better than Gaussian Process.