HW3

Aarshay Jain (aj2713)

Problem1: Gaussian Process Coding

a) Code to implement gaussian process

Code submitted separately for everything in a .py file

b) Test RMSE for parameter combinations

The results table is shown in Figure 1

		variance									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
b	5	1.96627564	1.93313496	1.92341988	1.92219731	1.92476887	1.92921232	1.93463387	1.94058292	1.94681977	1.95321235
	7	1.92016252	1.90487602	1.90807995	1.91590138	1.92480383	1.93370119	1.94225358	1.95037987	1.95809292	1.9654379
	9	1.89764785	1.90251864	1.91764736	1.93251408	1.94569928	1.95723463	1.96740312	1.97649163	1.98474067	1.99234112
	11	1.89050623	1.91498061	1.93884854	1.95793608	1.9732157	1.98576411	1.99637506	2.00560315	2.01383539	2.02134475
	13	1.89584777	1.93558563	1.96459718	1.9855019	2.00131421	2.01387841	2.02431035	2.03330676	2.04131748	2.04864154
	15	1.9096027	1.95954869	1.99080353	2.01191544	2.02737028	2.03946517	2.04946339	2.05810491	2.06584529	2.07297608

Figure 1: Test RMSE for combinations of b and σ^2

c) Best value of parameters

The parameters values $b = 11 \ and \ \sigma^2 = 0.1$ attained lowest test RMSE of 1.89. This lowest RMSE achieved in homework1 was around 2.2 so this model is performing better than a ridge regression with polynomial features.

Some drawbacks of this approach as compared to homework1 are:

- The closed form solution of a gaussian process is a computationally more expensive approach as we've to invert an nxn matrix in this as compared to ridge regression where we've to invert a dxd matrix. So if n grows large, gaussian process might become practically infeasible.
- Gaussian process can overfit significantly to the training set because it maps the input into an infinite dimensional space where a linearly separable plane is highly probable. So we need to run a careful grid-search on the model parameters. We need to run a search for the optimum value of regularization parameter in ridge regression as well, but its computationally easier to do when the data size is large.
- Feature selection is easier and intuitive in case of ridge regression as we can directly use the magnitude of feature coefficients and consider removing the one with low magnitude. Using the rbf kernel obviates calculation of weights and thus feature selection is infeasible in a gaussian process model.

d) Visualization using dimension 4

The required plot is shown in Figure 2

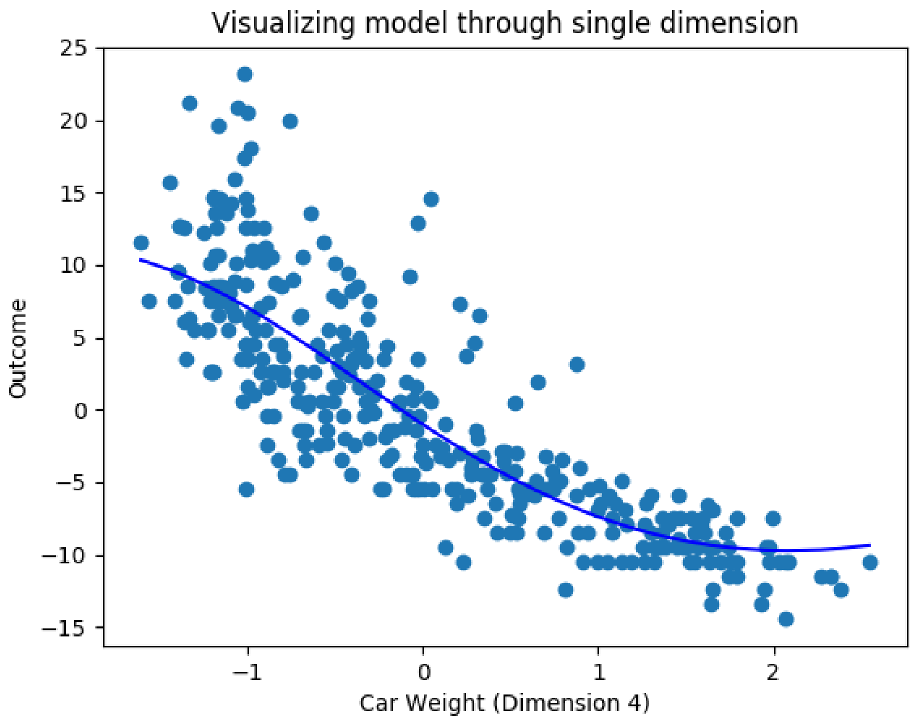

Figure 2: Scatter chart for dimension 4 with true and predicted values

Problem 2: Boosting Coding

a) Train and Test errors

The required plot is shown in Figure 3

Figure 3: Train and Test Errors vs Boosting Iteration

b) Upper Bound of Training Error

The required plot is shown in Figure 4

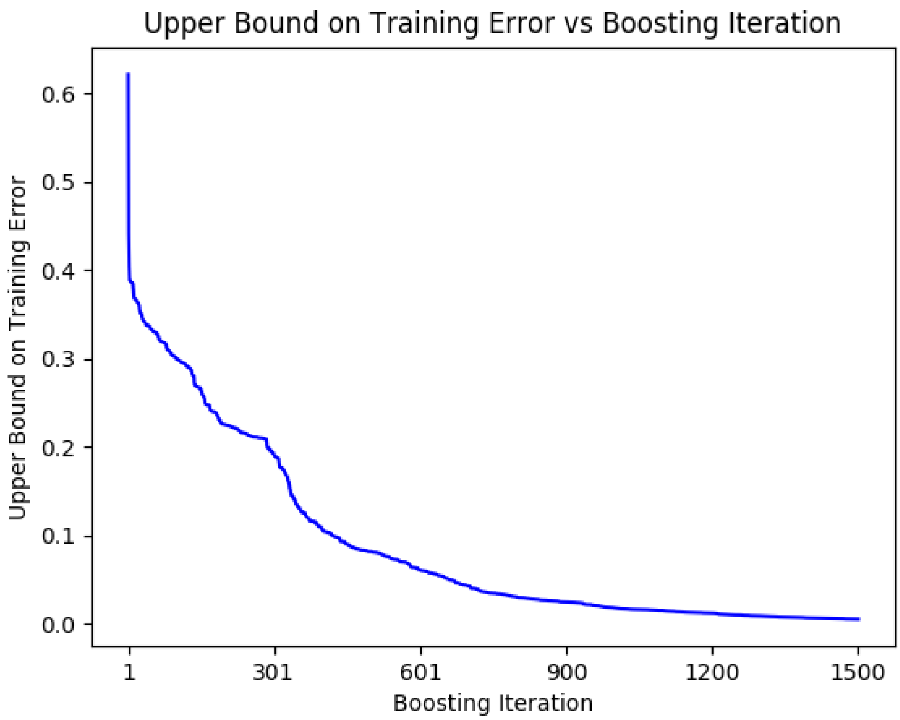

Figure 4: Upper Bound on Training Error vs Boosting Iteration

c) Histogram of occurrences of training data points

The required plot is shown in Figure 5

Figure 5: Total Number of selections of training data points

d) Variation of ϵ_t and α_t

The variation of ϵ_t is shown in Figure 6 while the variation of α_t is shown in Figure 7.

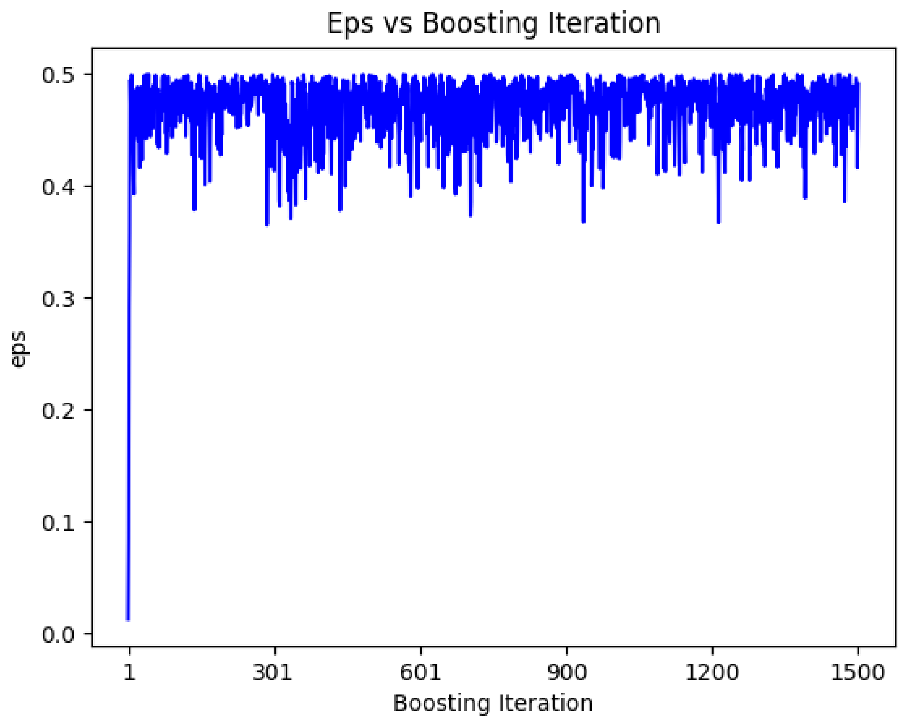

Figure 6: ϵ_t vs Boosting Iteration

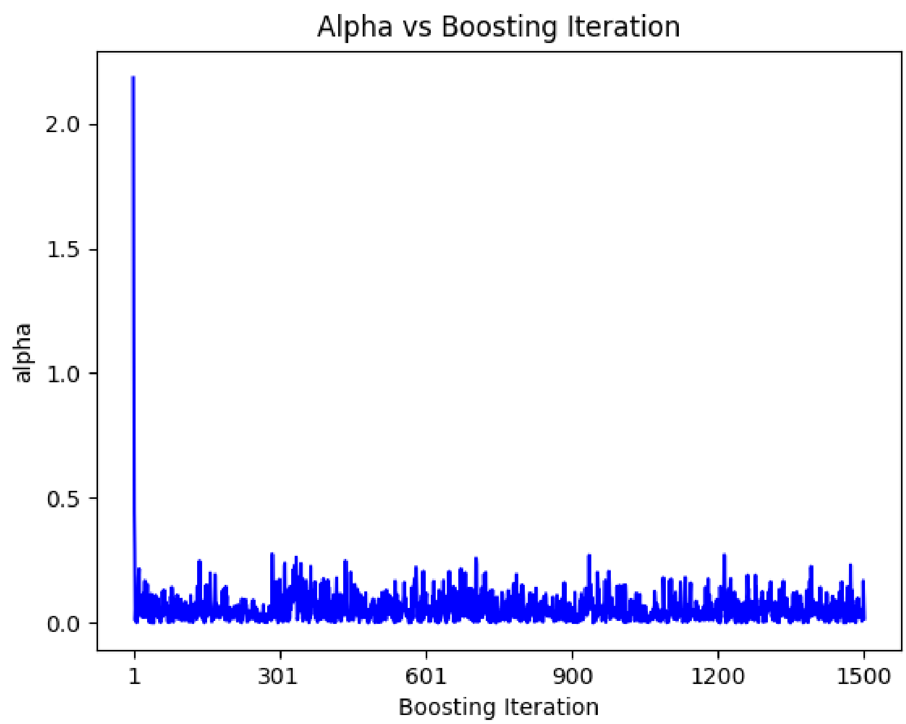

Figure 7: α_t vs Boosting Iteration