MATH 304 Project

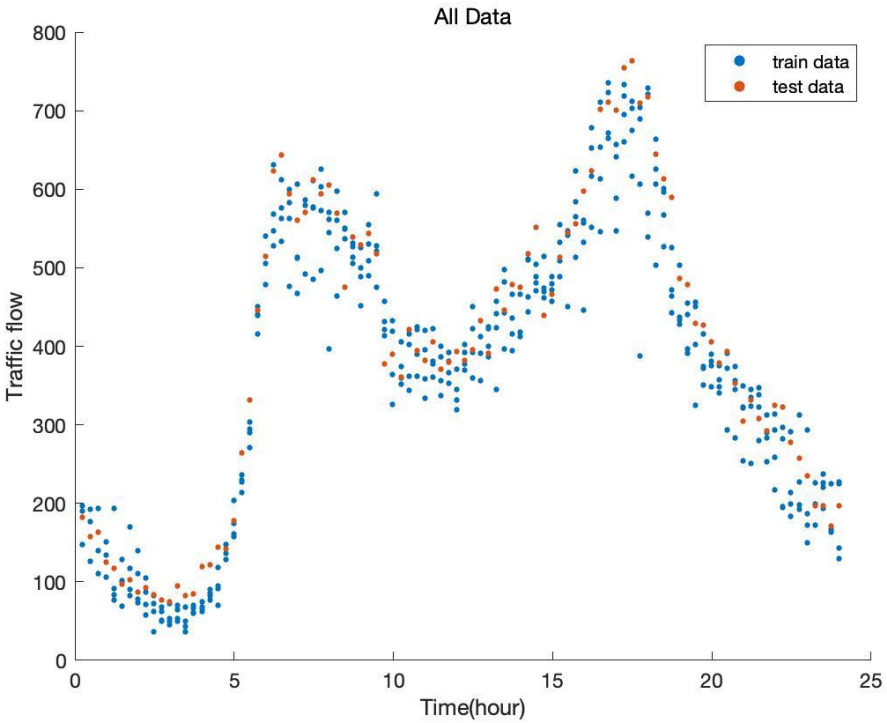
Traffic Flow Predictions with LSR and SVR

Haitong Lin
hl345@duke.edu

In this project, I use Least Square Regression (LSR) and Support Vector Regression (SVR) to predict the traffic flow information in June 2016.

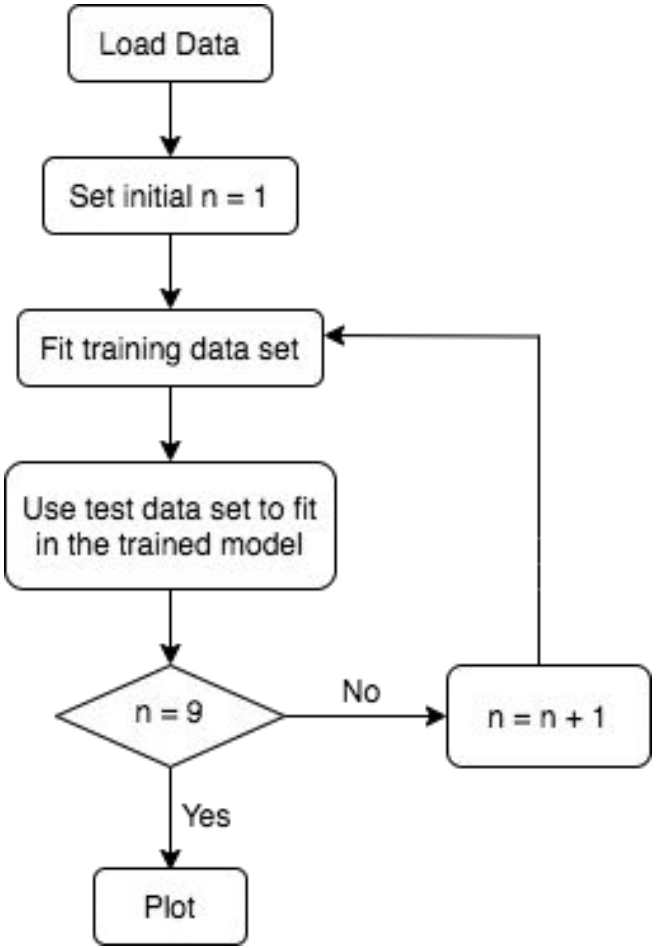
Introduction

Data Overview



Part 1: Least Square Regression (LSR)

The logic of LSR is to minimize the prediction error.



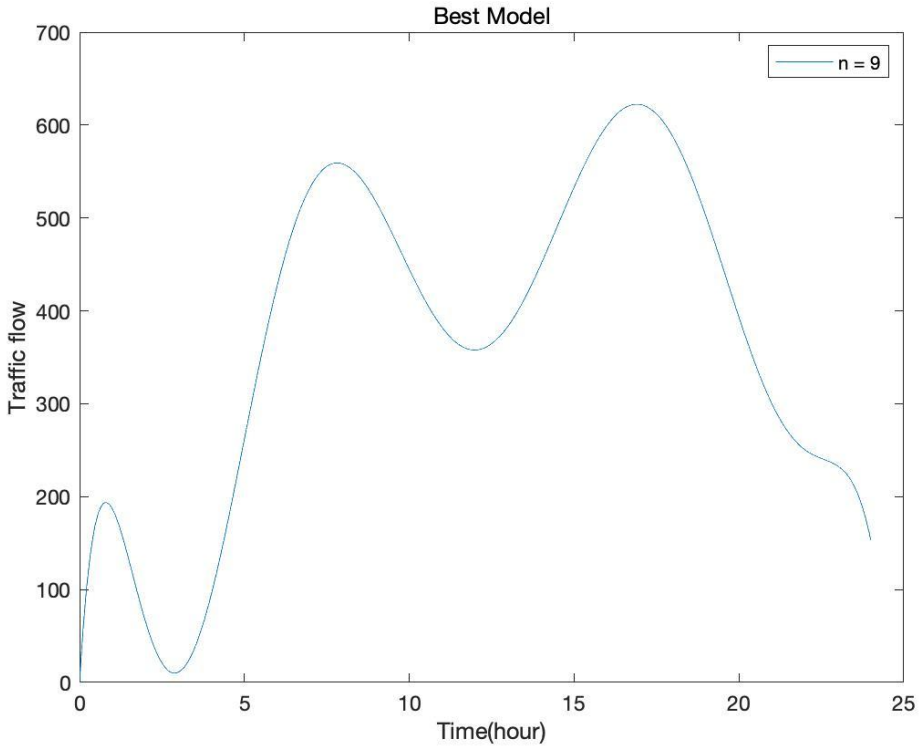
Part 1: LSR Results

Model		n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
Training error	MSE	29834	13195	12732	12687	11969	6315	6269.7	3080.2	2953.2
	r^2	0.1084	0.6057	0.6195	0.6208	0.6423	0.8113	0.8126	0.9079	0.9117
Test error	MSE	32158	15085	14313	14152	13453	6852.6	6816.2	3346.1	3329.7
	r^2	0.1305	0.6028	0.6237	0.6283	0.647	0.8285	0.8296	0.9258	0.9261

It is shown that as n grows, MSE decreases.

Part 1:
LSR Best Model

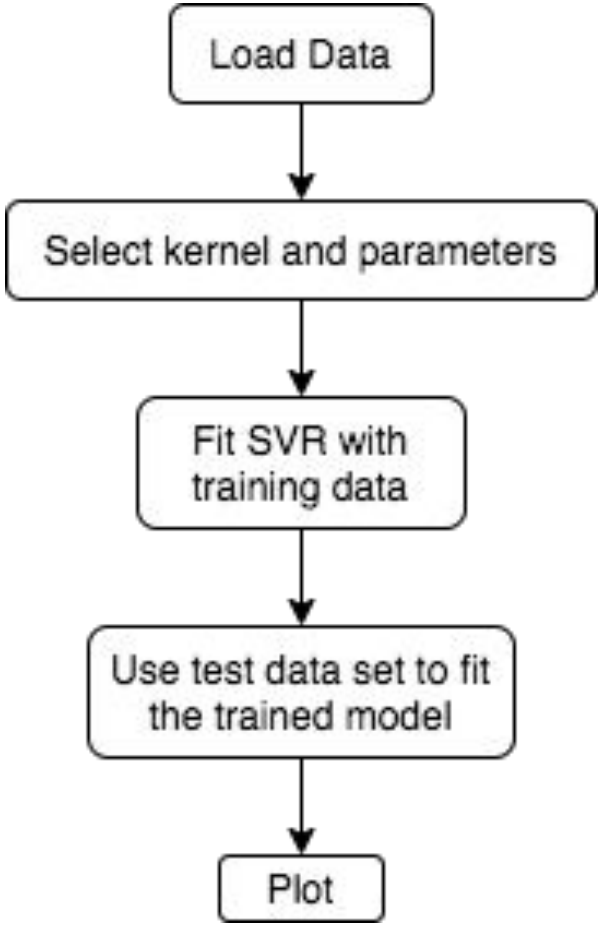
Among the different models we tested, n = 9 is the one with the least error for both training and testing data set.



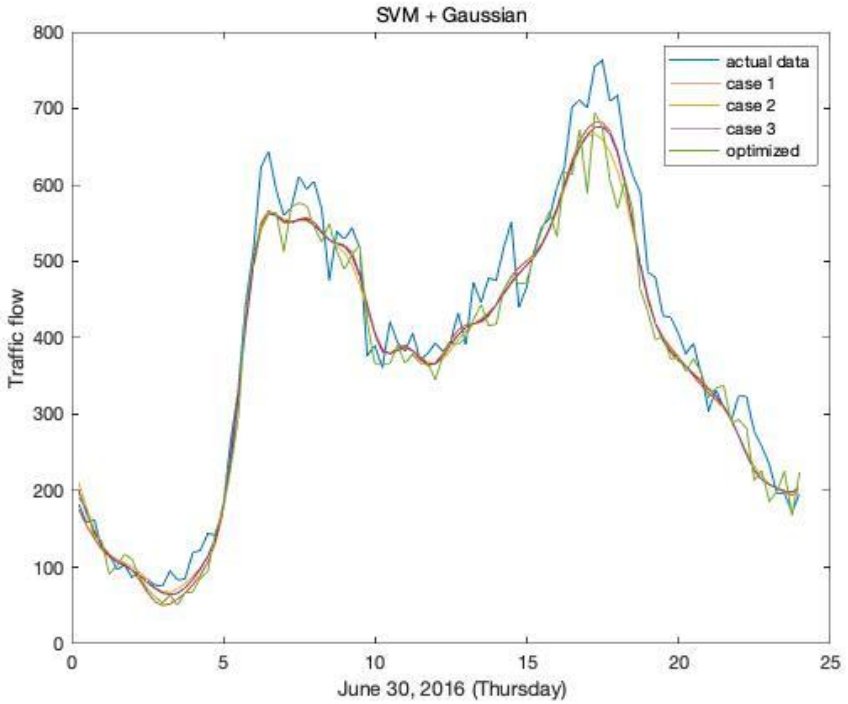
Part 2
Support Vector Regresion (SVR)

In SVR, we minimize the error while maximizing the margin, knowing that some errors are tolerated.

To solve nonlinear regression problems, we need to implement different kernels to SVR such that the original space can be mapped to a new space.



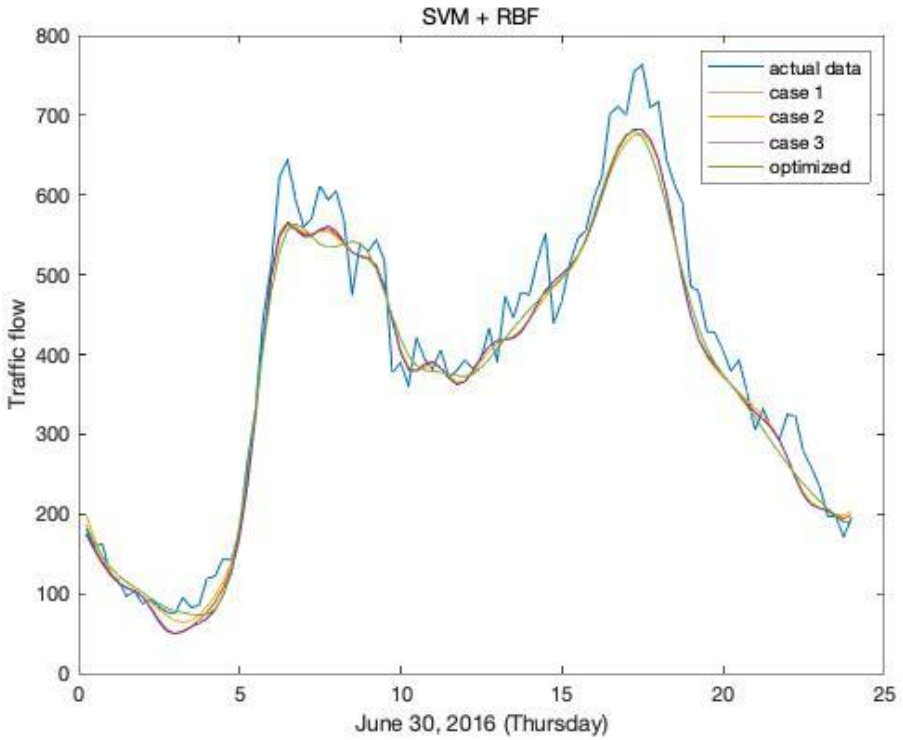
Part 2-1: SVR + Kernel 1 (Gaussian)



Model	SVM+Kernel 1(Gaussian)			
Setting	case 1	case 2 (50)	case 3 (100)	optimized
MSE	1426.8	1644.5	1466.4	2012.8
r^2	0.9771	0.9758	0.9772	0.9670

The case with the smallest error is case 1, which has default settings.

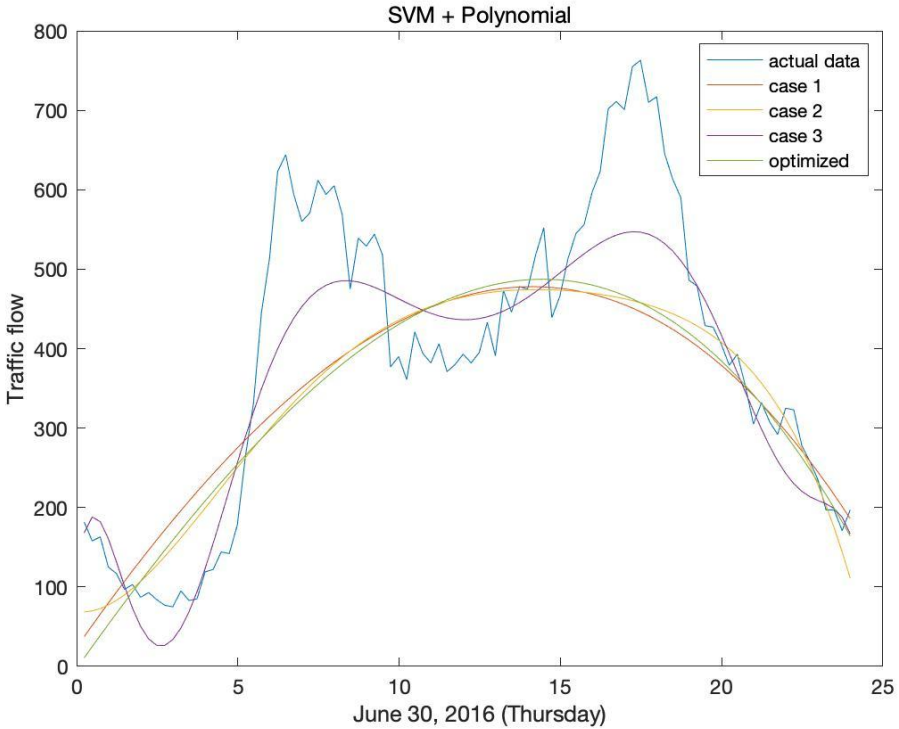
Part 2-2: SVR + Kernel 2 (RBF)



Model	SVM+Kernel 2 (RBF)			
Setting	case 1	case 2 (100)	case 3 (300)	optimized
MSE	1426.8	1466.4	1459.1	1565.5
r^2	0.9771	0.9772	0.9766	0.9740

The case with the smallest error is case 1, which has default settings.

Part 2-3: SVR + Kernel 3 (Polynomial)



Model	SVM+Kernel 3 (Polynomial)			
Setting	case 1	case 2 (5)	case 3 (10)	optimized
MSE	16898	15912	6630	16785
r^2	0.6155	0.6422	0.8565	0.6198

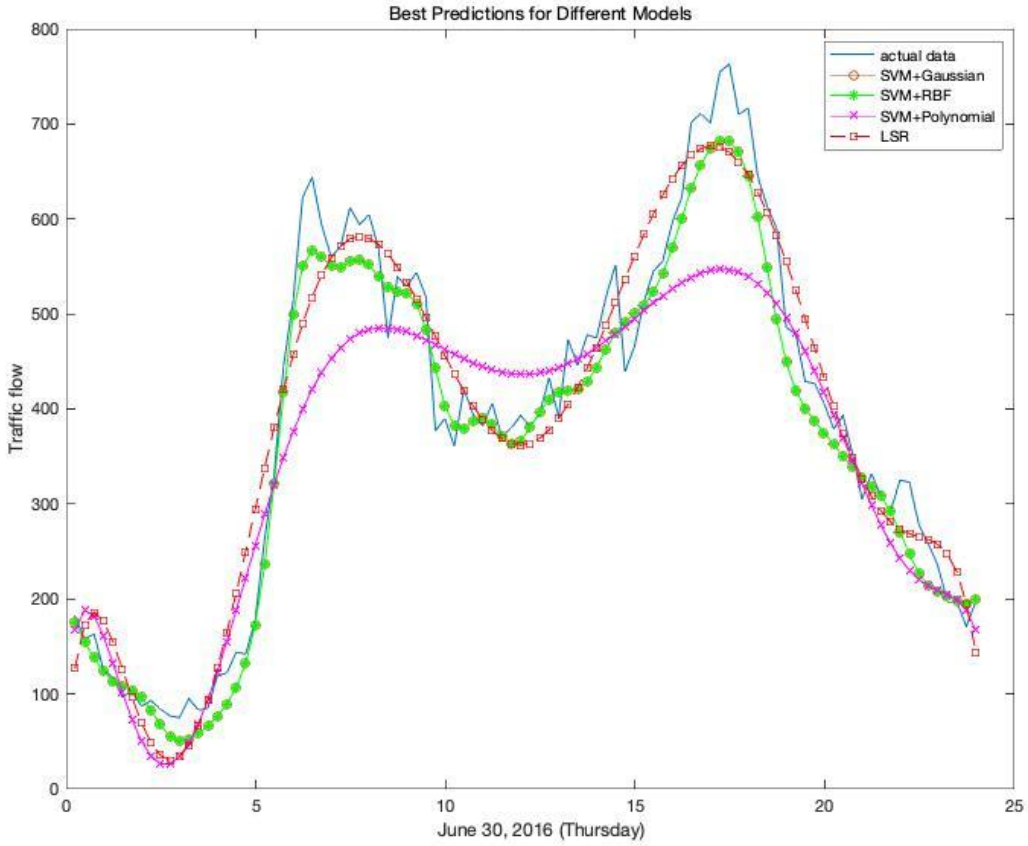
The case with the smallest error is case 3, which has a polynomial order of 10. Noted that the error for this model is huge.

Best Predictions For Each Model

Best models: SVR + Gaussian/RBF

Second best: LSR

Worst: SVR + Polynomial



Discussion

- Factors that affected accuracy
  - ➢ Data size
- Limits or problems
  - ➢ Number of models
  - ➢ Optimized cases
- Possible improvements
- The unique things that I did for this project
  - ➢ Model visualization

Thank you!