

---

# CS 419

# Assignment 1

---

Ashwin Kachhara  
10d070048

---

# Dataset (Computer Hardware)

---

- Regression
  - 10 attributes (6 predictive, 2 non-predictive, 1 goal field, 1 linear-regression's guess)
    - 1. vendor name: 30 (string)
    - 2. Model Name: many unique symbols (string)
    - 3. MYCT: machine cycle time in nanoseconds (integer)
    - 4. MMIN: minimum main memory in kilobytes (integer)
    - 5. MMAX: maximum main memory in kilobytes (integer)
    - 6. CACH: cache memory in kilobytes (integer)
    - 7. CHMIN: minimum channels in units (integer)
    - 8. CHMAX: maximum channels in units (integer)
    - 9. PRP: published relative performance (integer)
    - 10. ERP: estimated relative performance from the original article (integer)
-

# Dataset (contd)

---

	Min	Max	Mean	SD	PRP Corr
MYCT	17	1500	203.8	260.3	-0.3071
MMIN	64	32000	2868.0	3878.7	0.7949
MMAX	64	64000	11796.1	11726.6	0.8630
CACH	0	256	25.2	40.6	0.6626
CHMIN	0	52	4.7	6.8	0.6089
CHMAX	0	176	18.2	26.0	0.6052
PRP	6	1150	105.6	160.8	1.0000
ERP	15	1238	99.3	154.8	0.9665

# Problem

---

- Compare training and test error from different regressive models(LS regression, ridge regression, L1-support vector regression) on the computer hardware data
  - Predict PRP on the basis of values of MYCT, MMIN, MMAX, CACH, CHMIN, CHMAX
-

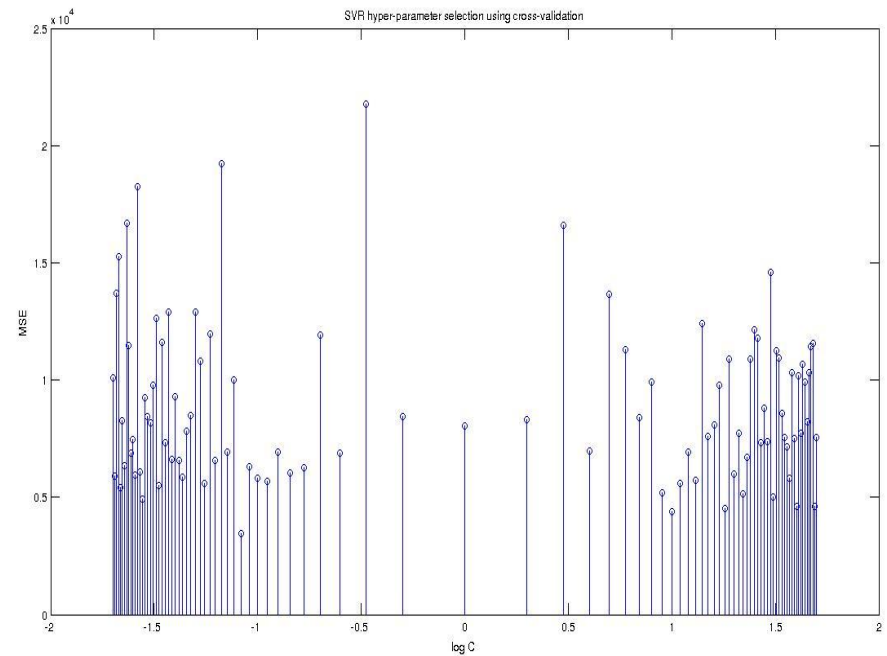
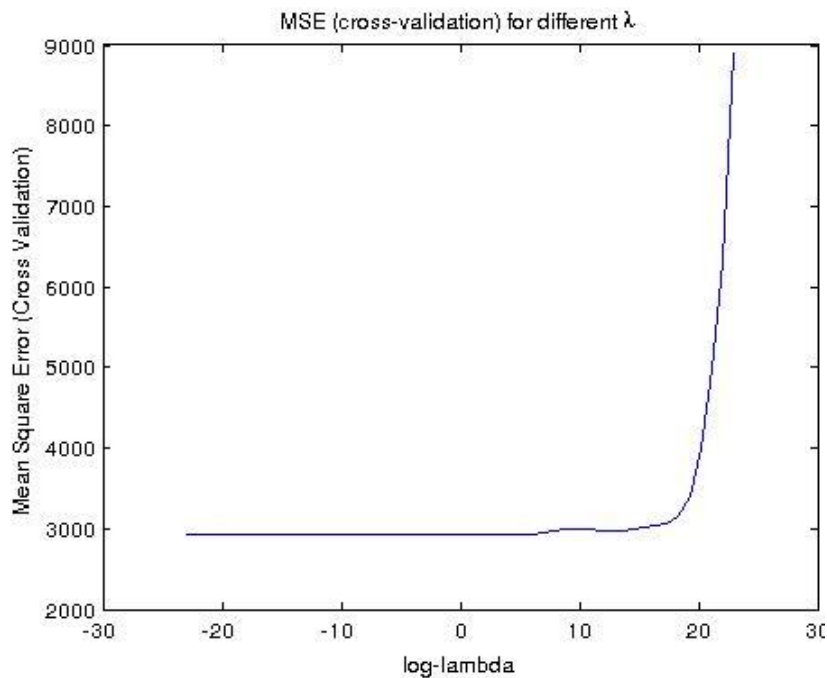
# Approach

---

- Read, organise data
  - Construct regressive model
    - Select hyper-parameter using 10-fold cross-validation (if applicable)
  - Compute training set mean square error
  - Compute test set mean square error
  - MATLAB - LS regression, Ridge regression
  - Python - L1-loss SVR
-

# Observations

	Training Set MSE	Test Set MSE
Least Squares Regression	1592.1	9390.7
Ridge Regression ( $\lambda = 175.7511$ )	958.66	13326
L1-loss SVR ( $C = 0.0833$ )	3459.16	57104.5



# Results

---

- There is overfitting in each of the models, as is evident from the fact that  $\text{testMse} > \text{trainMse}$
  - From the values in the previous slide, Least Square Regression model seems to be the optimum on the basis of Test and Train mse
-

---

Thank You!

---