**CS5402 Assignment 4 (Nonlinear Transform, Overfitting, Regularization, and Validation)**

**Due Nov 16 2016 11:00AM**

**1. Nonlinear transform and overfitting**

One data set () is provided in ‘RegressionData1.mat’, where matrix (N\*1) contains N training samples of dimension 1, i.e., , and (N\*1) is a column vector, i.e., .

Apply the Q-th order polynomial transform to the feature of each sample, (i.e., , then apply the linear regression algorithm to dataset to estimate .

(1) Choose Q=10, estimate , and compute the squared error cost:

Plot the N sample points along with the estimated curve (i.e., ).

(2) Choose Q =2 and repeat the above steps.

Compare the two results, what do you observe?

**2. Regularization**

The second data set () is provided in ‘RegressionData2.mat’, where matrix (N\*1) contains N training samples of dimension 1, i.e., , and (N\*1) is a column vector, i.e., .

Apply the Q-th order polynomial transform to the feature of each sample, (i.e., , then apply the regularized linear regression to dataset to estimate , i.e., obtain by solving .

(1) Choose Q=15 and , estimate , compute the squared error cost:

Plot the N sample points along with the estimated curve (i.e., ).

(2) Choose Q =15 and , repeat the above steps.

Compare the two results, what do you observe?

**3. Leave-one-out cross validation (optional, up to 10% bonus)**

The parameter is critical in the regularized linear regression. Apply the leave-one-out cross validation strategy to select the best from the pool of {0.00001, 0.0001, 0.001, 0.01, 0.1, 1}.

**4. Bonus to early birds (up to 10%)**: if you are the submission (), you earn (11-n)% bonus points for this homework.

Submit your codes and a written report to Canvas in which you briefly discuss at least the following things:

* Summarize what you think the homework was about (what was the task; what were you trying to achieve).
* Describe the algorithms you used to solve the homework problems.
* Run your program and show quantitative (error cost) and qualitative (pictures with sample points along with the regressed curves) results. Explain any design decisions you may have during the experiments.
* Summarize your observations from the experiments.