ELL409: Assignment 3

Maximum points: 6

Demo Schedule: 20th April 2016 (tentative)

10 April 2016

1. **Polynomial Curve Fitting**

   Polynomial curve fitting is an example of regression. Here you will apply the concepts of linear regression for polynomial curve fitting. In regression, the objective is to learn a function that maps an input variable x to a continuous target variable y.
   For this part, you will be provided a personalised input file that contains data of the form (x_i, y_i) for $i = 1, \ldots, 100$. The relationship between x and y is of the form:

   $$y = w_0 + w_1 x + \cdots + w_M x^M + \epsilon$$

   where the noise ϵ is drawn from a Gaussian distribution with zero mean and unknown (but fixed, for a given input file) variance. M is also unknown. You can download your input data file from `http://privateweb.iitd.ac.in/~seshan/a3/<groupno>` (for e.g., `http://privateweb.iitd.ac.in/~seshan/a3/group01`) where `groupno` is your group number as listed here: `http://web.iitd.ac.in/~seshan/ell409_assignment_groups.html`. The goal is to identify the underlying polynomial (both the degree and the coefficients), as well as obtain an estimate of the noise variance.
   Specifically, the following tasks are to be accomplished:

   - To begin with, use only the first 20 data points in your file. Solve the polynomial curve fitting regression problem using error function minimisation. Define your own error function other than the sum-of-squares error. Try different error formulations and report the results.
   - Use a goodness-of-fit measure for polynomials of different order. Can you distinguish overfitting, underfitting, and the best fit?
   - Obtain an estimate for the noise variance.
   - Introduce regularisation and observe the changes. For quadratic regularisation, can you obtain an estimate of the optimal value for the regularisation parameter λ? What is your corresponding best guess for the underlying polynomial? And the noise variance?
   - Now repeat all of the above using the full data set of 100 data points. How are your results affected by adding more data? Comment on the differences.
   - What is your final estimate of the underlying polynomial? Why?

   You will be required to give a demonstration of regression, the coefficients you have obtained, and how you have done so. In addition, present visualisations of the data and results in meaningful ways.

2. (Part-2 of this assignment will be updated in a couple of days)