Lecture Notes 1: Non-Linear Regression

Welcome to this course on *Non-Linear Regression*. I hope all of you have taken some courses in Statistics and Mathematics, and particularly on Linear Algebra, Mathematical Statistics and Linear Regression. Definitely, it will be really helpful if you have already taken these courses, otherwise you have to struggle a bit more.

First I would like to mention few important aspects of this course. As the name suggests it is an extension of the *Linear Regression* or *Linear Model* course. You may not find this course in most of the places. We have developed this course quite some time back at IIT Kanpur, and I have taught it also several times. But one thing I can tell that it has changed significantly along the time. Particularly today we find several applications in various Machine Learning techniques. I have incorporated several new topics and had to delete few to accommodate within this time frame. I am not going to follow one particular text book for this course. I will take help from the book by Seber and Wild on *Non-Linear Regression* and also by Bates and Watts in this same topic. I will be following several papers along the line.

I have already uploaded the first course handout, and you can see how the evaluations will be done. It will be based on four quizzes, one mid-sem, one-project and one final examinations. I have already announced the dates of the quizzes. They cannot be changed unless it is being declared as the National Holiday. In that case it will be on the next day of the class. The project might have a viva component, and that I will decide and announce after the mid-sem.

Now one natural question is what are main features of this course? Why do we need this course? What this course will offer? and what are the limitations of this course?

First let us discuss about the linear regression model. A simple linear regression model takes the following form:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

This is one of the oldest models which have been used to understand data.

It means there are two variables x_i and y_i and it is assumed that there is a linear relation between x_i and y_i. Moreover, y_i's are known as the dependent variables and x_i's are known as independent variables. Here β_0 is known as intercept, β_1 is the slope of the regression line, and ϵ_i is known as additive error. It is usually assumed that ϵ_i's are independent and identically distributed with mean zero and finite variance. The main problem is to find the *best* regression line, i.e. β_0 and β_1, and also the properties of these estimators. Now to develop the properties of these estimators we need the suitable assumptions on the error components.

A more general multiple linear regression model takes the following form:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \epsilon_i.$$

In this case instead of one independent variables, we have p independent variables and it is assumed that the dependent variable y_i has a linear relation with all these independent variables. Again in this case also under suitable assumptions on the additive error components ϵ_i, the main issue is to fine suitable estimators of the unknown linear parameters namely $\beta_0, \beta_1, \ldots, \beta_p$ and find the properties of these estimators. We have a very well developed theory and estimation procedure in case of linear and multiple regressions models, and we will be describing briefly later.

If there does not exist any linear relation between the dependent and independent variables, then clearly the linear regression is not going to work. Let us look at the following example:

Data on the amount of heat generated by friction were obtained by Count Rumford in 1798. A bore was fitted into a stationary cylinder and pressed against the bottom by means of a screw. The bore was turned by a team of horses for 30 minutes and after that the temperature was measured at small intervals of time. The following data have been obtained, see Table 1 We plot the data in Figure 1. It is not linear. Try to fit a linear regression line, and see what will be the best fitted line.

Let us look at another example. See for example the gold price data of 1gm pure gold in Rupees in the Indian market for two months (Figure 2). Clearly, the gold price does not follow a linear pattern with time. Therefore,

Table 1: Temperature versus time for Rumford cooling experiment

Time (min)	Temperature (^{o}F)	Time (min)	Temperature (^{o}F)
4	126	24	115
5	125	28	114
7	123	31	113
12	120	34	112
14	119	37.5	111
16	118	41	110
20	116		

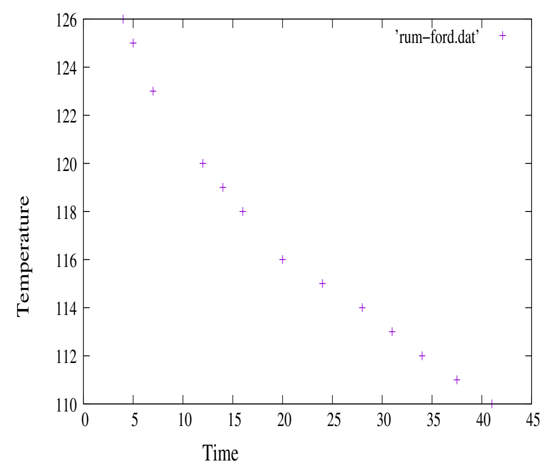

Figure 1: Scatter plot of Rumford's cooling data

a linear model (linear regression) cannot be used to explain the gold price in terms of time. We need some non-linear function explain the behavior of the gold price with respect to time. So the main question is how to analyze these kind of data sets.

In this course that will be the main topic of interest. We will develop several non-linear models to analyze these kind of data sets. The main purpose of this course is (i) what kind of models we need, (ii) how do we estimate the parameters, (iii) what are the assumptions we need (iv) what are the properties of these estimators etc. In this course we will be dealing with several

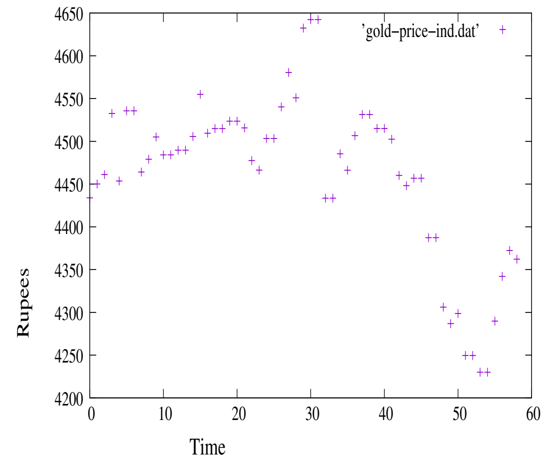

Figure 2: Gold price data for two months in the Indian market.

models which are of the general form:

$$y = f(\mathbf{x}, \theta) + \epsilon. \tag{1}$$

Here y is the dependent variables, $\mathbf{x}$ is the independent vector valued (may be) variable, $f(\cdot)$ is a non-linear function and ϵ is the error random variable. Throughout this course it will be assumed that the function $f(\cdot)$ is known, we observe y and $\mathbf{x}$ and the parameter θ is unknown.

Observe that in case of linear regression model

$$f(\mathbf{x}, \theta) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

Therefore, whatever methodologies and theories we are going to develop they should be applicable in case of linear regression model also.

The following questions we would like to address

1. Based on the observations from the model (1) estimate the unknown parameter θ under a suitable assumption on the error component ϵ.
2. Develop the properties of these estimators?
3. Check the model validity, i.e. whether the given model (1) works for a given data set or not.