1 Non-linear regression

In epidemiology studies, a common focus is on assessing changes in a response distribution with a continuous exposure, adjusting for covariates. For example, Longnecker in 2001 studied the relationship between the DDT metabolite DDE and preterm delivery. The substance DDT is effective against malaria-transmitting mosquitoes, and so is widely used in malaria-endemic areas in spite of growing evidence of health risks. The Longnecker et al. (2001) study measured DDE in mothers serum during the third trimester of pregnancy, while also recording the gestational age at delivery, GAD, and demographic factors, such as age. Following standard practice in reproductive epidemiology, Longnecker et al. (2001) dichotomized GAD using a 37-week cut-off, so that deliveries occurring prior to 37 weeks of completed gestation were classified as preterm. In this project, the goal is to fit two non-linear regression models

1. Using the raw gestational ages and using cubic polynomial in each variable as the regressor

2. Dichotomizing the gestational age at delivery as Longnecker did and then fit a logistic regression in R using cubic polynomial in each variable as the regressor

Divide the dataset into training and test and find out which one of these two models give better prediction.