CS181 Homework 1

Ruitao (Toby) Du

02/06/2015

1 Non-Uniformly Weighted Data

First, we put r_n inside the quadratic form.

$$\begin{aligned} E_D(\boldsymbol{w}) &= \frac{1}{2}\sum_{n=1}^{N} r_n\{t_n - \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_n)\}^2 \\ &= \frac{1}{2}\sum_{n=1}^{N}\{\sqrt{r_n}t_n - \boldsymbol{w}^T\sqrt{r_n}\boldsymbol{\phi}(\boldsymbol{x}_n)\}^2 \end{aligned} \tag{1.1}$$

We assume x_n is D dimensions. We can write equation (1.1) in matrix form.

$$E_D(\boldsymbol{w}) = \frac{1}{2}(\boldsymbol{\Phi}w - T)^T(\boldsymbol{\Phi}w - T) \tag{1.2}$$

where $\boldsymbol{\Phi}$ is an N $\times$D matrix, whose elements are given by $\boldsymbol{\Phi}_{n,j} = \boldsymbol{\phi}_j(\boldsymbol{x}_n)$, so that

$$\boldsymbol{\Phi} = \begin{bmatrix} \sqrt{r_1}\boldsymbol{\phi}_1(\boldsymbol{x}_1) & \sqrt{r_1}\boldsymbol{\phi}_2(\boldsymbol{x}_1) & \cdots & \sqrt{r_1}\boldsymbol{\phi}_D(\boldsymbol{x}_1) \\ \sqrt{r_2}\boldsymbol{\phi}_1(\boldsymbol{x}_2) & \sqrt{r_2}\boldsymbol{\phi}_2(\boldsymbol{x}_2) & \cdots & \sqrt{r_2}\boldsymbol{\phi}_D(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{r_N}\boldsymbol{\phi}_1(\boldsymbol{x}_N) & \sqrt{r_N}\boldsymbol{\phi}_2(\boldsymbol{x}_N) & \cdots & \sqrt{r_N}\boldsymbol{\phi}_N(\boldsymbol{x}_N) \end{bmatrix}, T = \begin{bmatrix} \sqrt{r_1}t_1 \\ \sqrt{r_2}t_2 \\ \vdots \\ \sqrt{r_N}t_N \end{bmatrix}$$

And then we can take the gradient of w and set this equation to 0.

$$\bigtriangledown_w E_D(\boldsymbol{w}) = \boldsymbol{\Phi}^T(\boldsymbol{\Phi}w - T) = 0 \quad \Rightarrow \quad \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{w} = \boldsymbol{\Phi}^T T$$

We can caclculate the solution w^* that minimize the function

$$w^* = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T T$$

2 Priors and Regularization

Since$p(\boldsymbol{w}|\alpha) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha^{-1}\boldsymbol{I})$ and $p(\boldsymbol{t}|\boldsymbol{w}) = \prod_{n=1}^{N} \mathcal{N}(t_n|\boldsymbol{w}^T\boldsymbol{\Phi}(x_n), \beta^{-1})$, we know that

$$p(\boldsymbol{w}|\alpha) = \frac{1}{\sqrt{(2\pi)^D|\alpha^{-1}\boldsymbol{I}|}} exp(-\frac{1}{2}\boldsymbol{w}^T\alpha\boldsymbol{w})$$

$$p(\boldsymbol{t}|\boldsymbol{w}) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\beta^{-1}} exp(-\frac{1}{2}\beta(t_n - \boldsymbol{w}^T\boldsymbol{\Phi}(x_n))^2)$$

so we can have following equation

$$\begin{aligned}\ln p(\boldsymbol{w}|\boldsymbol{t}) &= \ln p(\boldsymbol{w}|\alpha) + \ln p(\boldsymbol{t}|\boldsymbol{w}) \\ &= -\frac{1}{2}\boldsymbol{w}^T\alpha\boldsymbol{w} - \frac{1}{2}\sum_{n=1}^{N}\beta(t_n - \boldsymbol{w}^T\boldsymbol{\Phi}(x_n))^2 + const \\ &= -\frac{\beta}{2}\sum_{n=1}^{N}\{t_n - \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_n)\}^2 - \frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w} + const\end{aligned} \tag{2.1}$$

Take $\lambda = \alpha/\beta$ and drop constant terms. We can write equation (2.1) in following form.

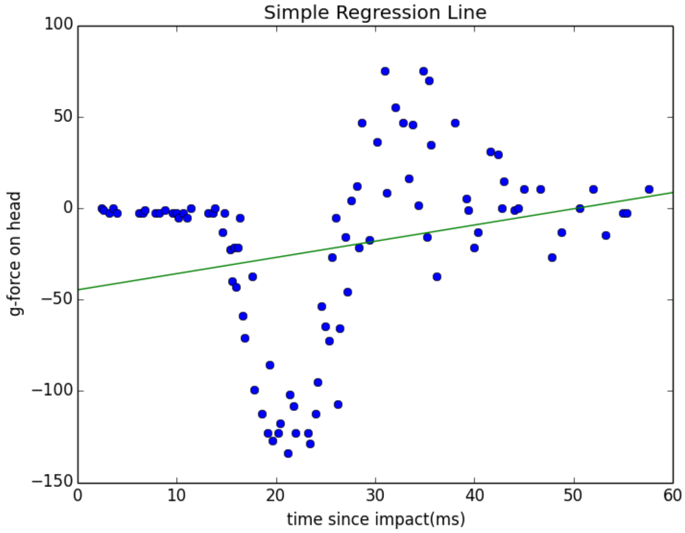
$$\begin{aligned}\ln p(\boldsymbol{w}|\boldsymbol{t}) &= -\beta(\frac{1}{2}\sum_{n=1}^{N}\{t_n - \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_n)\}^2 + \frac{\alpha}{2\beta}\boldsymbol{w}^T\boldsymbol{w}) + const \\ &= -(\frac{1}{2}\sum_{n=1}^{N}\{t_n - \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_n)\}^2 + \frac{\lambda}{2}\boldsymbol{w}^T\boldsymbol{w})\end{aligned} \tag{2.2}$$

According to the question, we can write $\ln p(\boldsymbol{w}|\boldsymbol{t})$ as a function of $E_D(\boldsymbol{w})$ and $E_W(\boldsymbol{w})$,

$$\ln p(\boldsymbol{w}|\boldsymbol{t}) = -(E_D(\boldsymbol{w}) + \lambda E_W(\boldsymbol{w})) \tag{2.3}$$

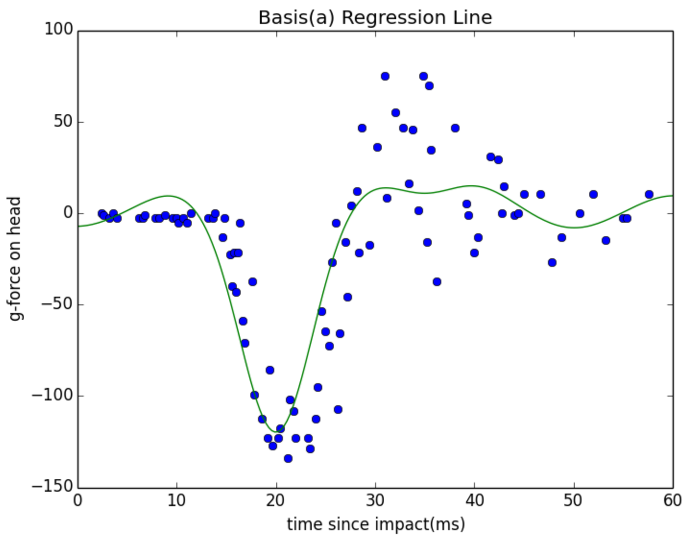
Therefore, maximizing the log posterior is equivalent to minimizing the regularized error term given by $E_D(\boldsymbol{w}) + \lambda E_W(\boldsymbol{w})$

3 Modeling Motorcyce Helmet Forces



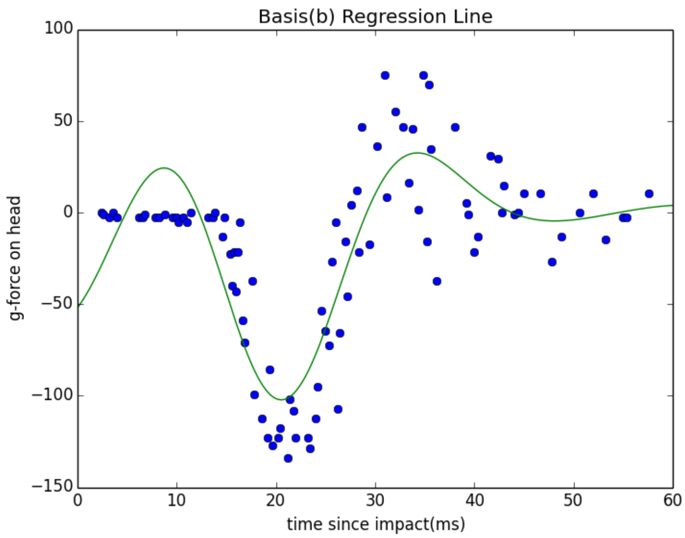
It is underfitting. It does not follow the trend of the data because simple linear regression assumes linearity, which is not fulfilled in this data.

In the following basis functions, I do not add $\phi_0(x) = 1$ and just follow the basis functions provided by the questions.
(a) $\phi_j(x) = exp\{-((x - 10j)/5)^2\}$ for $j = 0, ..., 6$



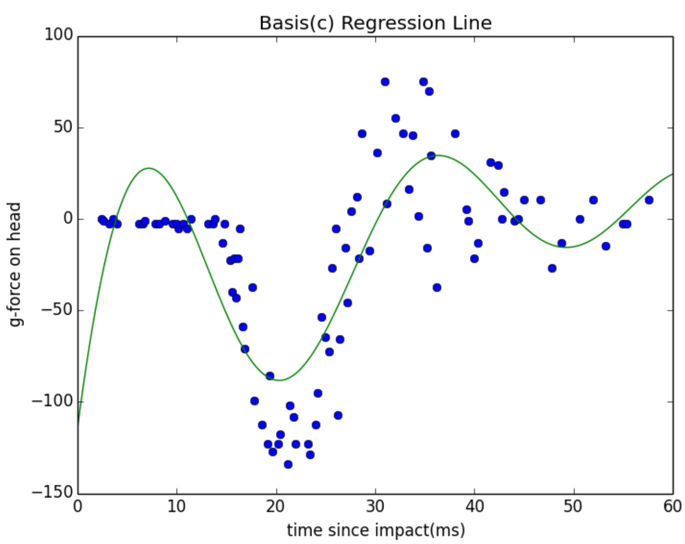
It fits the data well compared to other basis functions. It captures the change of the data except a little error at the beginning.

(b) $\phi_j(x) = exp\{-((x-10j)/10)^2\}$ for $j = 0, ..., 6$



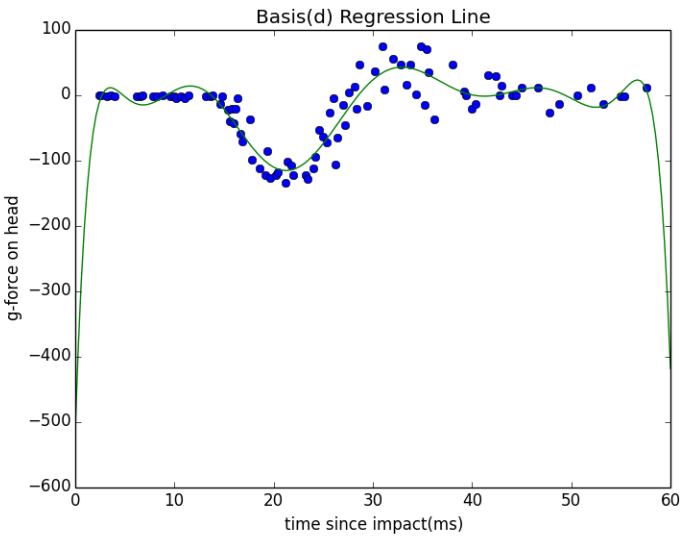
It slightly overfits the data. From the left side of the plot, we can see that the regression line is parabola but the data does not change.

(c) $\phi_j(x) = exp\{-((x-10j)/25)^2\}$ for $j = 0, ..., 6$

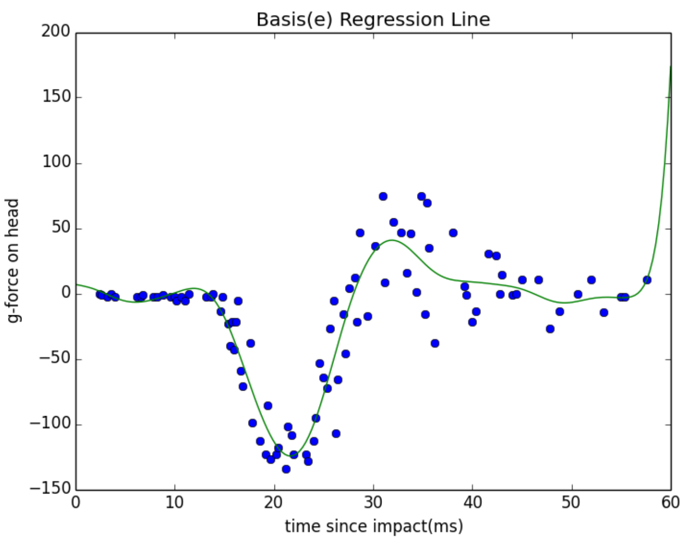


It overfits the data. There is a curve at the beginning of the plot. From the left side of the plot, we can see that the regression line is parabola but the data does not change.

(d) $\phi_j(x) = x^j$ for $j = 0, ..., 10$



It overfits the data. Even though it follows the trend of the data, when input value goes out of the range of the original data, the dependent variable changes substantially, which is not realistic.
(e) $\phi_j(x) = sin\{x/j\}$ for $j = 1, ..., 20$



It overfits the data. When independent variable goes greater, the dependent variable suddenly rises substantially. Similar to basis function (d), this basis function overfits the data.

4 Calibration

It took me about 5 hours to finish it.