Homework 1: Linear Regression

Writeup due 23:59 on Friday 6 February 2015

You will do this assignment individually and submit your answers as a PDF via the Canvas course website. There is a mathematical component and a programming component to this homework.

1. Non-Uniformly Weighted Data [7pts]

Consider a data set in which each data point t_n is associated with a weighting factor $r_n > 0$, so that the sum-of-squares error function becomes

$$E_D(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N} r_n\{t_n - \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_n)\}^2.$$

Find an expression for the solution $\boldsymbol{w}^*$ that minimizes this error function.

2. Priors and Regularization [7pts]

Consider the Bayesian linear regression model given in Bishop 3.3.1. The prior is

$$p(\boldsymbol{w} \mid \alpha) = \mathcal{N}(\boldsymbol{w} \mid \boldsymbol{0}, \alpha^{-1}\boldsymbol{I}),$$

where α is the precision parameter that controls the variance of the Gaussian prior. The likelihood can be written as

$$p(\boldsymbol{t} \mid \boldsymbol{w}) = \prod_{n=1}^{N} \mathcal{N}(t_n \mid \boldsymbol{w}^\mathsf{T}\boldsymbol{\phi}(\boldsymbol{x}_n), \beta^{-1}),$$

Using the fact that the posterior is the product of the prior and the likelihood (up to a normalization constant), show that maximizing the log posterior (i.e., $\ln p(\boldsymbol{w} \mid \boldsymbol{t}) = \ln p(\boldsymbol{w}|\alpha) + \ln p(\boldsymbol{t} \mid \boldsymbol{w}$ is equivalent to minimizing the regularized error term given by $E_D(\boldsymbol{w}) + \lambda E_W(\boldsymbol{w})$ with

$$E_D(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}(t_n - \boldsymbol{w}^\mathsf{T}\boldsymbol{\phi}(\boldsymbol{x}_n))^2$$

$$E_W(\boldsymbol{w}) = \frac{\lambda}{2}\boldsymbol{w}^\mathsf{T}\boldsymbol{w}$$

Do this by writing $\ln p(\boldsymbol{w} \mid \boldsymbol{t})$ as a function of $E_D(\boldsymbol{w})$ and $E_W(\boldsymbol{w})$, dropping constant terms if necessary. Conclude that maximizing this posterior is equivalent to minimizing the regularized error term given by $E_D(\boldsymbol{w}) + \lambda E_W(\boldsymbol{w})$. (Hint: take $\lambda = \alpha/\beta$)

3. Modeling Motorcycle Helmet Forces [10pts]

The objective of this problem is to learn about linear regression with basis functions by modeling the g-forces associated with motorcycle helmet impacts. Download the file `motorcycle.csv` from the course website. It has two columns. The first one is the number of milliseconds since impact and the second is the g-force on the head. The data file looks like this:

```
"time since impact (ms)","g force"
2.4,0
2.6,-1.3
3.2,-2.7
3.6,0
4,-2.7
6.2,-2.7
6.6,-2.7
6.8,-1.3
...
```

and you can see a plot of the data in Figure 1.

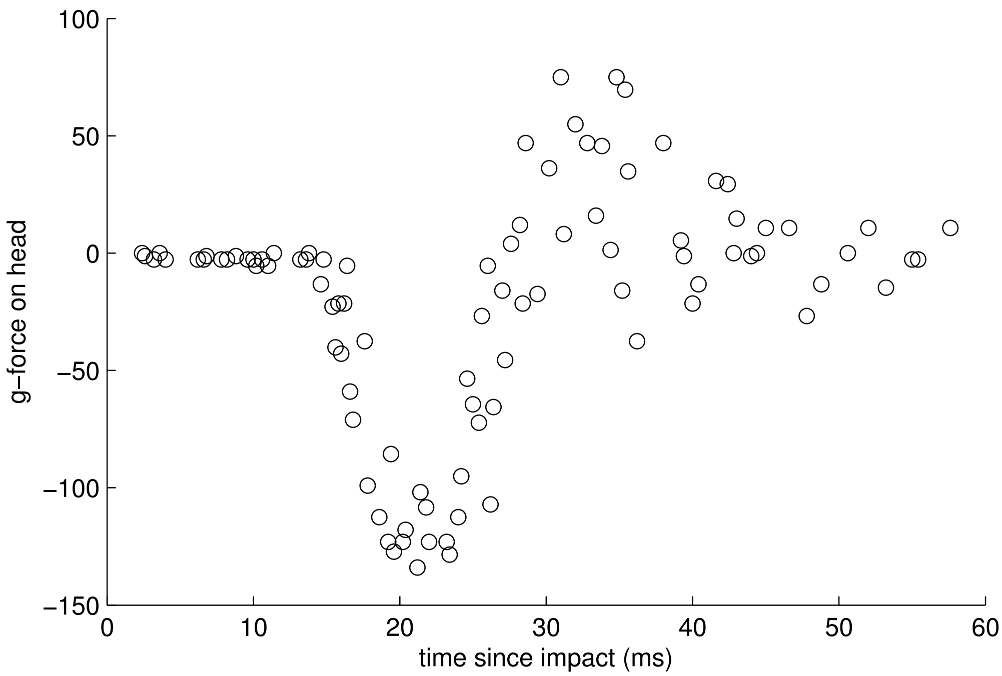

Figure 1: Motorcycle crash helmet data. The horizontal axis is time since impact and the vertical axis is force on the head.

Implement basis function regression with ordinary least squares.[1] Some sample Python code is provided in `linreg.py`. Plot the data and regression lines for the simple linear case, and for each of the following sets of basis functions:

(a) $\phi_j(x) = \exp\{-((x - 10j)/5)^2\}$ for $j = 0, \ldots, 6$

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

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

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

(e) $\phi_j(x) = \sin\{x/j\}$ for $j = 1, \ldots, 20$

In addition to the plots, provide one or two sentences for each, explaining whether you think it is fitting well, overfitting or underfitting. If it does not fit well, provide a sentence explaining why.

Calibration [1pt]

Approximately how long did this homework take you to complete?

Changelog

- **v1.0** – 28 January 2015 at 13:00

[1] Note that the data clearly don't have fixed variance! There is obviously less variance on the left of the plot. Modeling such *heteroscedastic* data is beyond the scope of the course.