Homework 3 - Statistical modelling and inference

October 29, 2015

Students	**Student IDs**
Max van Esso	73539
Marco Fayet	125593
Felix Gutmann	125604

1 Homework lecture 5

1.1 Excercise 1

To show: $f(\mu) = (\mu - a) + \lambda|\mu|$ is minimised at $f(a - \frac{\lambda}{2})$ for $\mu > 0$

Starting from the equation, which is by definition a piecewise function:

$$f(\mu) = (\mu - a) + \lambda|\mu| = \begin{cases} (\mu - a) + \lambda\mu \\ (\mu - a) - \lambda\mu \end{cases}$$

Since we should find the minimum in the positive area, we ignore the negative part and take the derivative of the postive side and set it to zero:

$$\begin{aligned} \frac{\partial f(\mu)}{\partial \mu} &= 0 \\ 2(\mu - a) + \lambda &= 0 \\ 2(\mu - a) &= -\lambda \\ \mu &= (-\frac{\lambda}{2} + a) \end{aligned}$$

Take the second derivative to indentify the character of the extremum:

$$\frac{\partial^2 f(\mu)}{\partial \mu^2} = 2$$

The second derivative is greater than zero which implies a **minimum**.

1.2 Excercise 2

To show: Closed form of $\boldsymbol{w}_{MAP}$ for $\boldsymbol{t}_n \sim N(\boldsymbol{w}, \boldsymbol{q}^{-1}\boldsymbol{I})$ with prior $p(\boldsymbol{w}) \propto exp\{-(\delta/2)\} \sum_i |w_i|$

Prerquisite remarks:

- The letter c denotes a constant (see line 5 and 6)

Excercise:

$$\begin{aligned}
\boldsymbol{w}_{MAP} &= \arg\max \log p(w|t) \\
&= \arg\max \log p(t|w) + \log p(w) \\
&= \arg\max \log \prod_n p(t_n|w) + \log \exp\{-(\delta/2)\sum_i |w_i|\} \\
&= \arg\max \sum_n \log N(w_n, q^{-1}\boldsymbol{I}) + \log \exp\{-(\delta/2)\sum_i |w_i|\} \\
&= \arg\max \sum_n c \log \exp\{-\frac{1}{2}q(t_n - w)^T(t - w)\} + \log \exp\{-(\delta/2)\sum_i |w_i|\} \\
&= \arg\max \sum_n -\frac{c}{2}q(t_n - w)^T(t_n - w) - (\delta/2)\sum_i |w_i| \\
&\propto \min \sum_n q(t_n - w)^T(t_n - w) + \delta \sum_i |w_i| \\
&= \min q \sum_n \sum_i (t_{ni} - w_i)^2 + \delta \sum_i |w_i|
\end{aligned}$$

Minimise that function by setting the deriviative to zero:

$$\begin{aligned}
\frac{\partial}{\partial w} &= 0 \\
&= -2q\sum_n (t_{ni} - w_i) + \delta \frac{w_i}{|w_i|} = 0
\end{aligned}$$

Note that we will only consider the positive side of the absolute part. Therefore $\frac{w_i}{|w_i|}$ simplyfies to (+)1 and the equation becomes:

$$= -2q\sum_n (t_{ni} - w_i) + \delta = 0$$

$$\begin{aligned}
\sum_n (t_{ni} - w_i) &= \frac{\delta}{2}q^{-1} \\
-Nw_i + \sum_n t_{ni} &= \frac{\delta}{2}q^{-1} \\
Nw_i + \frac{\delta}{2}q^{-1} &= \sum_n t_{ni} \\
w_i &= (\sum_n t_{ni} - \frac{\delta}{2}q^{-1})\frac{1}{N}
\end{aligned}$$

1.3 Excercises in R

Ancillary remarks on results (see page 4 - 6):

1.3.1 Learn g from the data

At first we want to learn g from the data. Note that we also have a second parameter δ. Both g and δ are not independent from each other and interact somehow (see lecture slides).

Procedure to find stable g:

1. Fix a value for δ
2. Define a starting value for g
3. Compute the predictive mean using those inputs
4. Compute the precision of the residuals
5. Start again from step 2 using this new precision
6. Break in case of a stable value

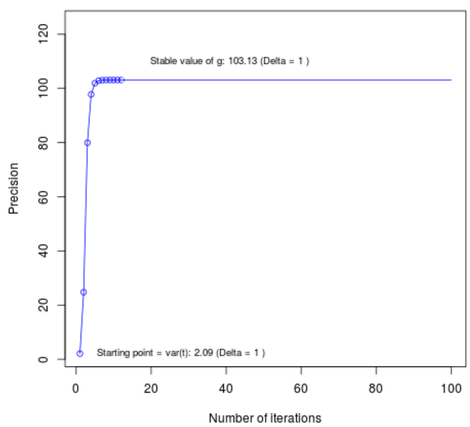
Results:

- Since we are starting from the pure data as the only information, the first idea was to start with the precision of t (implying predictive mean of zero)
- Quick developement towards a stable value ($g \approx 103.13, for\ \delta = 1$) after approx. 13 iterations (c.f. Figure 1 (a),(b), (c))
- Starting point seem irrelevant based on two abitrary chosen starting points (c.f. Figure 1 (b), (c))
- For increasing δ the stable value of g decreases (c.f. Figure 1 (d))

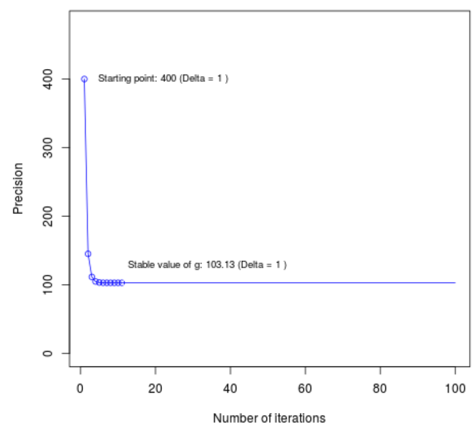
1.3.2 Sensitivity of δ

Results:

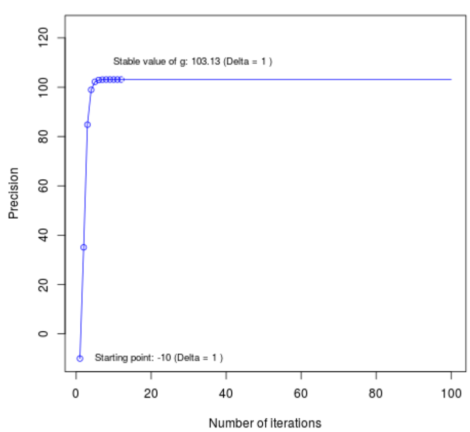
- Low values for δ leads towards overfitting (c.f Figure 2 (a),(b), (c),(d))
- Low values for δ result in a broader range of posterior draws (c.f Figure 3 (a))
- Increasing δ result in closer posterior draws (c.f Figure 3 (a), (b), (c),(d))



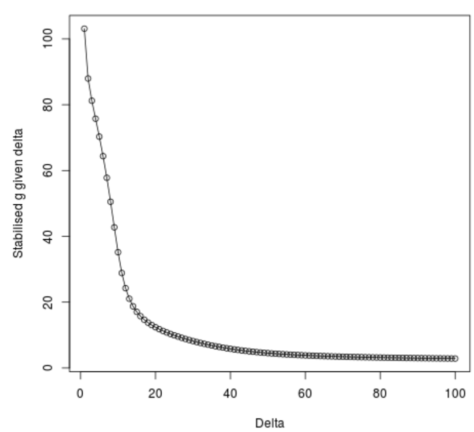
(a) Stable g starting from precision of t ($\delta = 1$)



(b) Stable g, abitrary start +400 ($\delta = 1$)

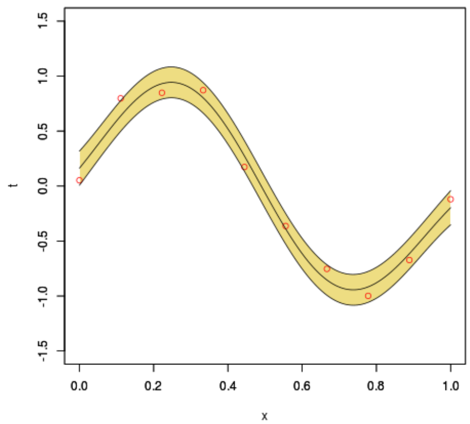


(c) Stable g, abitrary start -10 ($\delta = 1$)

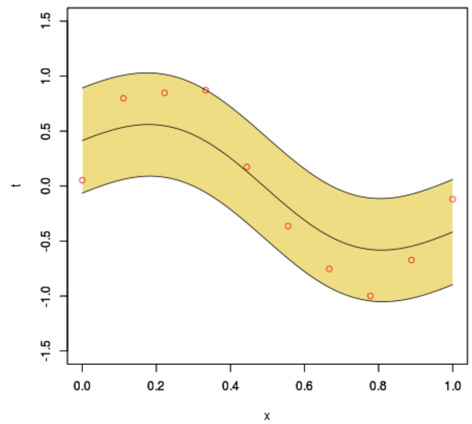


(d) Stable gs, given dif. delta

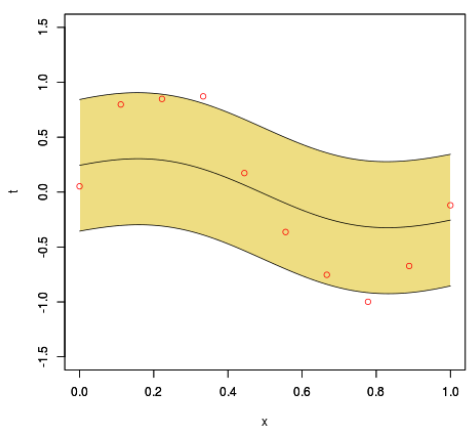
Figure 1: Find stable value for g



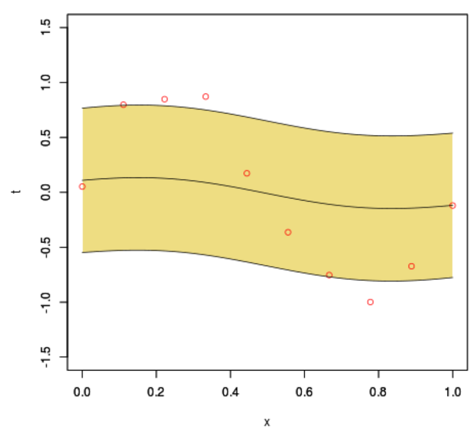
(a) Predictive mean with one σ ($\delta = 1$)



(b) Predictive mean with one σ ($\delta = 25$)

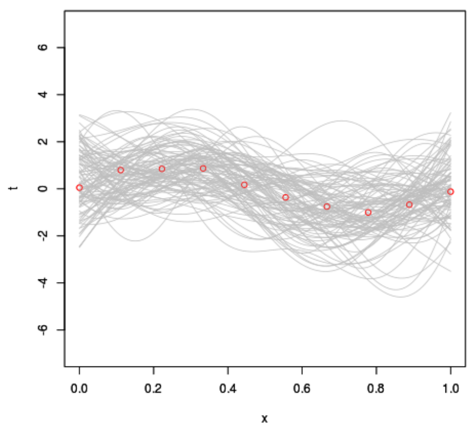


(c) Predictive mean with one σ ($\delta = 50$)

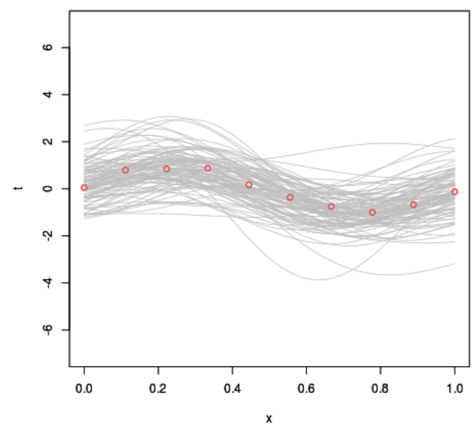


(d) Predictive mean with one σ ($\delta = 100$)

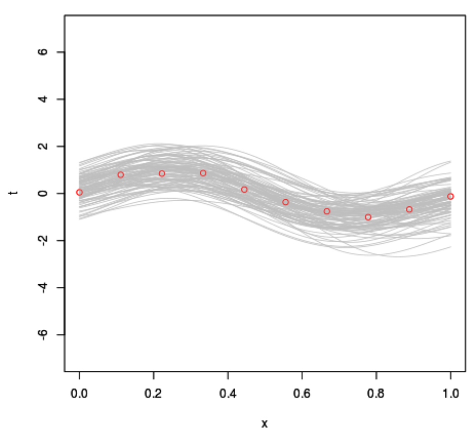
Figure 2: Different level of *delta*



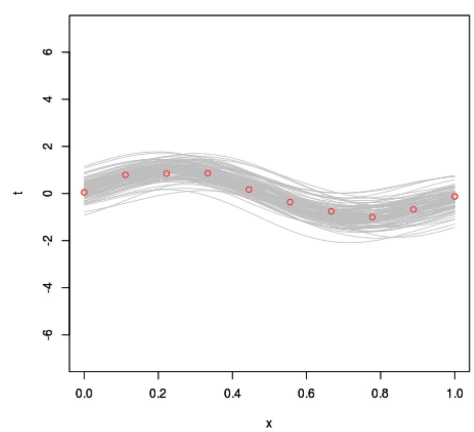
(a) Posterior draws ($\delta = 1$)



(b) Posterior draws ($\delta = 25$)



(c) Posterior draws ($\delta = 50$)



(d) Posterior draws ($\delta = 100$)

Figure 3: Posterior draws for different levels of δ