Barcelona Graduate School of Economics Master's Degree in Data Science

Statistical Modeling and Inference – Problem Set #6

Group 3: Niti Mishra · Miquel Torrens · Bálint Ván

November 30$^{\text{th}}$, 2015

Solution to proposed exercises.

Exercise 1

Part (a)

Given $t_n \sim \text{Laplace}(\mu, \Lambda)$, the likelihood function is:

$$\mathcal{L}(\text{Laplace}(t_n|\mu,\Lambda)) = \prod_{n=1}^{N} f(t_n|\mu,\Lambda) = \left(\frac{\Lambda}{2}\right)^N \exp\left\{-\Lambda \sum_{n=1}^{N} |t_n - \mu|\right\}$$

Part (b)

If we take the MLE:

$$\max_{\mu,\Lambda} \log \mathcal{L} = \max_{\mu,\Lambda} N \log \Lambda - \Lambda \sum_{n=1}^{N} |t_n - \mu| + C$$

With C containing the terms not dependent on μ or Λ. Then:

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \mu} = 0 \quad &\Leftrightarrow \quad -\Lambda \sum_{n=1}^{N} \frac{t_n - \mu}{|t_n - \mu|} = 0 \\ &\Leftrightarrow \quad \sum_{n=1}^{N} \text{sgn}(t_n - \mu_{MLE}) = 0 \qquad (1) \end{aligned}$$

Note that with odd N, for (1) to hold, $(N-1)/2$ observations need to have $t_n \geq \mu$ and the other $(N-1)/2$ have $t_n \leq \mu$. Thus:

$$\mathbb{E}[t] = \mu_{MLE} = \text{median}(t_1, t_2, ..., t_n)$$

is the expected value and also the median.

Part (c)

We derive Λ_{MLE}:

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \Lambda} = 0 \quad &\Leftrightarrow \quad N\frac{1}{\Lambda} - \sum_{n=1}^{N} |t_n - \mu| = 0 \\ &\Leftrightarrow \quad \Lambda_{MLE} = \left(\frac{1}{N}\sum_{n=1}^{N} |t_n - \mu|\right)^{-1} \end{aligned}$$

Part (d)

By the equivariance property[1] of the MLE:

$$\mathrm{var}[t] = \Sigma_{MLE} = \frac{2}{\Lambda_{MLE}^2} = 2\left(\frac{1}{N}\sum_{n=1}^{N}|t_n - \mu|\right)^2 .$$

[1] Wasserman's Theorem 10.14.

Exercise 2

Part (a)

Given:

$$t_n \sim \mathcal{N}(t_n|\mathbf{x}_n, \mathbf{w}, \eta_n, q) = \frac{(\eta_n q)^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{\eta_n q}{2}(t_n - \phi(\mathbf{x}_n)\mathbf{w})^T(t_n - \phi(\mathbf{x}_n)\mathbf{w})\right\}$$

and:

$$\eta_n \sim \text{Gam}\left(\eta_n|\frac{\nu}{2}, \frac{\nu}{2} - 1\right) = \frac{\left(\frac{\nu}{2} - 1\right)^{\nu/2}}{\left(\frac{\nu}{2} - 1\right)!}\eta_n^{\left(\frac{\nu}{2}-1\right)} \exp\left\{-\left(\frac{\nu}{2} - 1\right)\eta_n\right\}$$

We know:

$$\begin{aligned}
p(\eta_n|t_n, \mathbf{x}_n, \mathbf{w}, q) &\propto p(t_n|\eta_n, \mathbf{x}_n, \mathbf{w}, q)p(\eta_n) \\
&= C\eta_n^{\left(\frac{\nu}{2}-1+\frac{1}{2}\right)} \exp\left\{-\left(\frac{\nu}{2} - 1\right)\eta_n - \frac{\eta_n q}{2}\mathbf{e}_n^T\mathbf{e}_n\right\} \\
&= C\eta_n^{\left(\frac{\nu+1}{2}-1\right)} \exp\left\{\left(-\left(\frac{\nu}{2} - 1\right) - \frac{q}{2}\mathbf{e}_n^T\mathbf{e}_n\right)\eta_n\right\} \\
&= C\eta_n^{\left(\frac{\nu+1}{2}-1\right)} \exp\left\{-\left(\frac{\nu + q\mathbf{e}_n^T\mathbf{e}_n}{2} - 1\right)\eta_n\right\}
\end{aligned}$$

Where C contains the multiplication of the terms of the density functions not dependent on η_n. Observe this is a Gamma distribution with parameters:

$$\begin{aligned}
\alpha_n &= \frac{\nu + 1}{2} \\
\beta_n &= \frac{\nu + q\mathbf{e}_n^T\mathbf{e}_n}{2} - 1
\end{aligned}$$

This finishes the proof.

Part (b)

Given[2]:

$$Q(\theta, \theta') = \int \log\left(p(t, \eta|\theta)\right) p(\eta|t, \theta')d\eta = \mathbb{E}\left[\log\left(p(t, \eta|\theta)\right)\right]$$

[2]Equations (4.3) and (4.4) in Omiros' notes on the appendix of the lecture.

We can use $\theta = (q, \mathbf{w})$ and through maximum likelihood compute:

$$
\begin{aligned}
\mathbb{E}\left[\log p(t, \eta|\theta)\right] &= \mathbb{E}\left[\log\left(p(t|\theta)p(\eta|t, \theta')\right)\right] \\
&= \mathbb{E}\left[\log p(t|\theta) + \log p(\eta|t, \theta')\right] \\
&= \mathbb{E}\left[\log\left(\prod_{n=1}^{N}\left(\frac{q\eta_n}{2\pi}\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}qe_n^2\eta_n\right\}\right) + \log p(\eta|t, \theta')\right] \\
&= \mathbb{E}\left[\log\left(\prod_{n=1}^{N}\left(\frac{q\eta_n}{2\pi}\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}qe_n^2\eta_n\right\}\right)\right] + \mathbb{E}\left[\log p(\eta|t, \theta')\right] \\
&= \mathbb{E}\left[\log\left(\left(\frac{q\eta_n}{2\pi}\right)^{\frac{N}{2}}\exp\left\{-\frac{1}{2}q\sum_{n=1}^{N}e_n^2\eta_n\right\}\right)\right] + c \\
&= \mathbb{E}\left[\frac{N}{2}\log q - \frac{q}{2}\sum_{n=1}^{N}(t_n - \phi(\mathbf{x}_n)\mathbf{w})^2\eta_n\right] + c \\
&= \frac{N}{2}\log q - \frac{q}{2}(\mathbf{t} - \mathbf{\Phi}\mathbf{w})^T\mathbb{E}\left[\mathbf{H}\right](\mathbf{t} - \mathbf{\Phi}\mathbf{w}) + c
\end{aligned}
$$

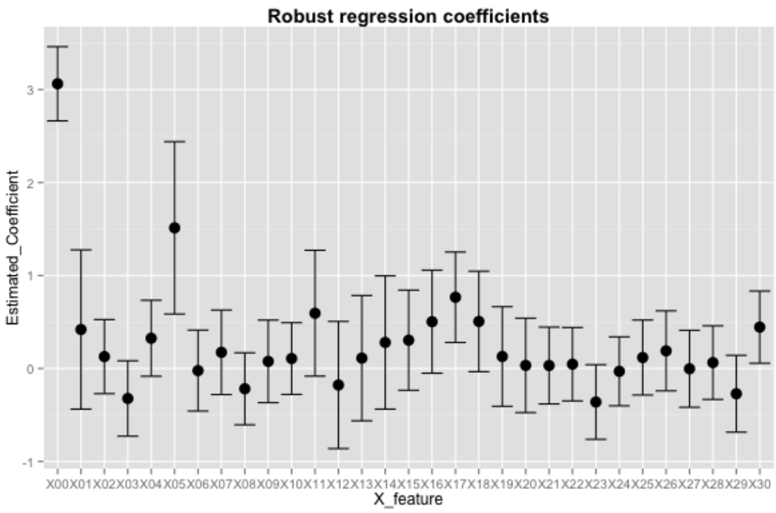
Where $\mathbf{H}$ is a diagonal matrix with $\mathbf{H}_{nn} = \eta_n$, and c contains all terms not dependent on θ. The distribution of η_n will not depend on θ but on the previous $\theta' = (q', \mathbf{w}')$, and given that for $t \sim \mathrm{Gam}(\alpha, \beta)$ we have $\mathbb{E}[t] = \alpha/\beta$, then:

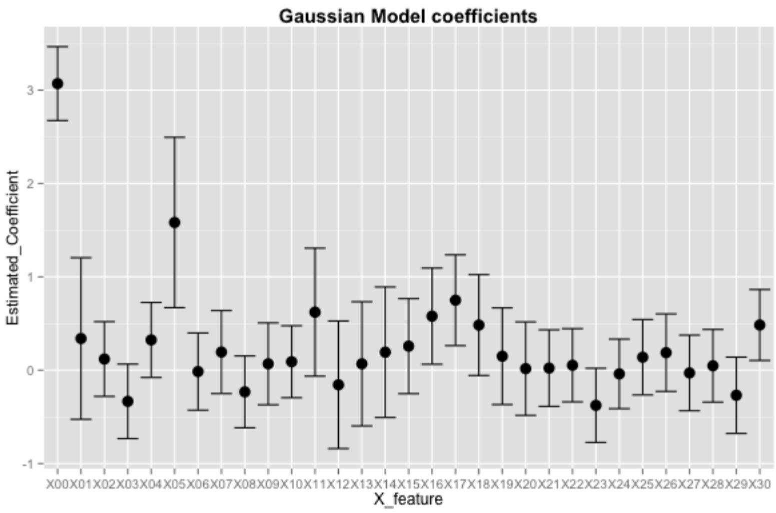
$$
\mathbb{E}[\eta_n] = \frac{\frac{\nu+1}{2}}{\frac{\nu+q'(e_n')^2}{2} - 1} = \frac{\nu+1}{\nu+q'(e_n')^2 - 2} = \frac{\nu+1}{\nu+q'(t_n - \phi(\mathbf{x}_n)\mathbf{w}')^2 - 2}
$$

Hence proved.

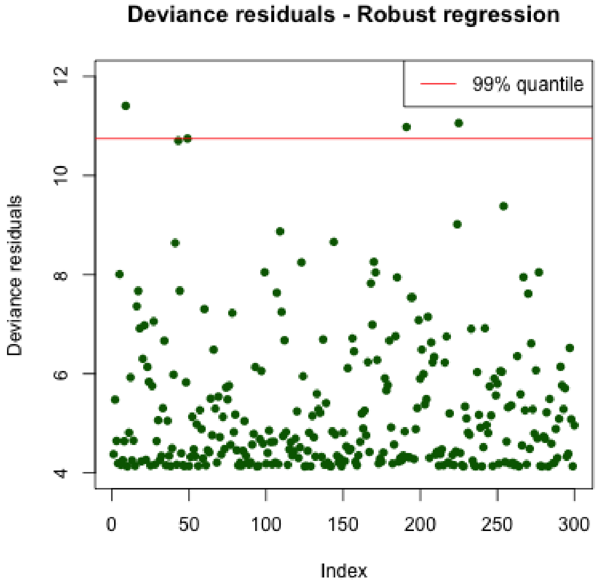
Exercise 3

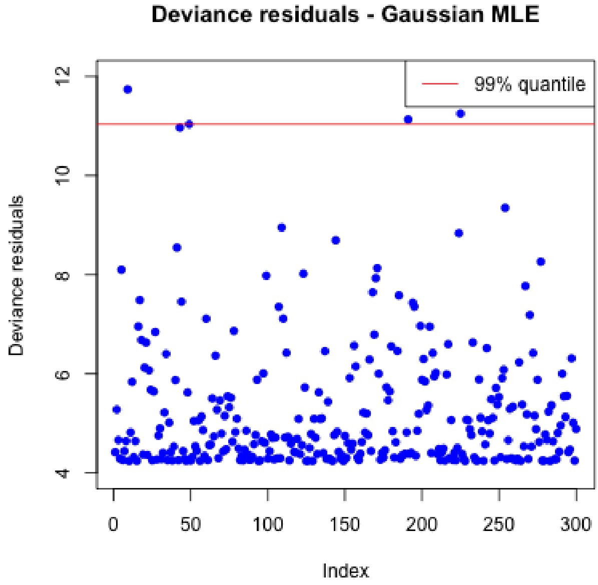
Part (a)





Part (b)

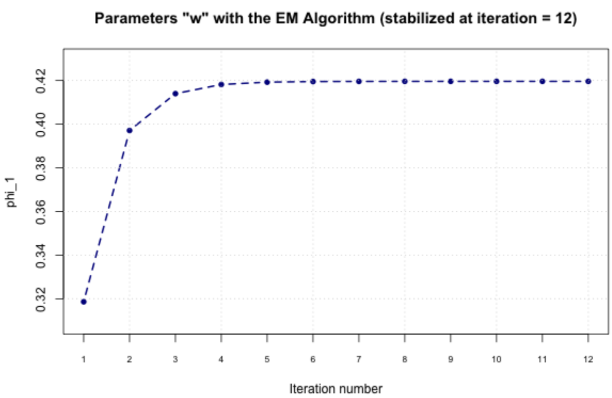


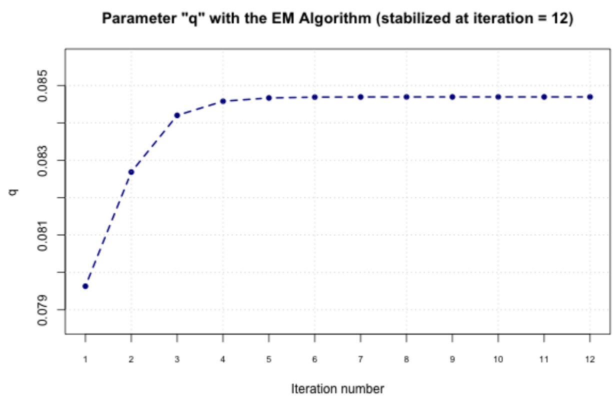


Part (c)

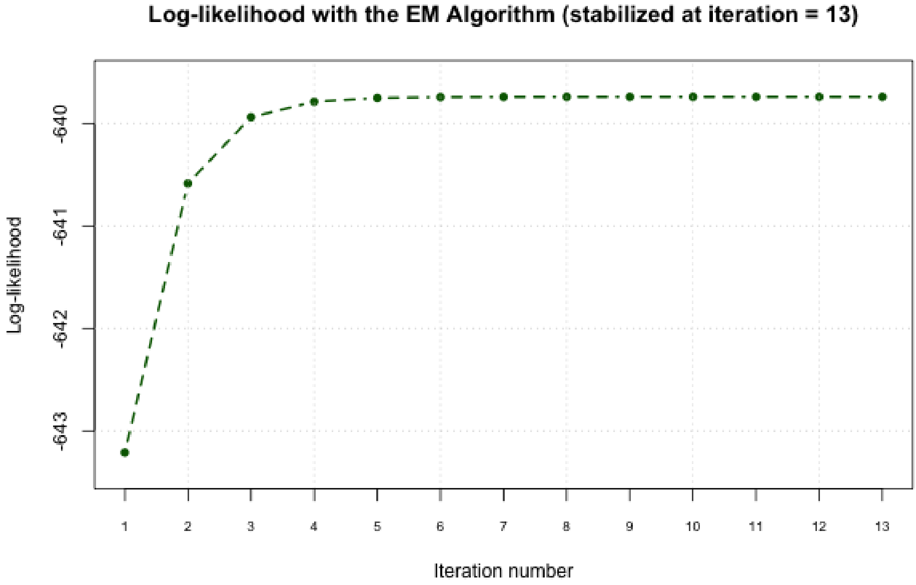
We can decide to stop the EM algorithm when the values for the estimated parameters θ estabilize, which is very similar to saying when the log-likelihood stops increasing (the sequential iterations increase its value by a very small number, say 10^{-6}).

Example (parameter stabilization: q and $\mathbf{w}$, here only w_1 for illustration):





Example (log-likelihood stabilization):



Part (d)

We can choose ν by running the EM algorithm for different values of ν and then choosing the smallest ν that maximizes the log-likelihood, i.e. the one for which increasing ν raises log-likelihood by a very small amount, again say by 10^{-6}.

Graphical representation: