Weighted Least Squares

Rahul Ramachandran
cs21btech11049

Rishit D
cs21btech11053

October 1, 2023

Contents

1 Solution **1**
1.1 Likelihood and Prior 1
1.2 ML and MAP Estimation 1
1.3 Error Function and Solution 2

1 Solution

1.1 Likelihood and Prior

In a homoscedastic setting, the likelihood is given by:

$$p(t_n|x_n, \mathbf{w}) \sim \mathcal{N}(t_n|\mathbf{w}^T\phi(x_n), \beta^{-1})$$

We arrive at this by assuming that the noise is Gaussian and independent of the input:

$$t_n = y(x_n, \mathbf{w}) + \epsilon$$

In a heteroscedastic setting, we instead assume that the noise depends on the input. In particular, we will assume that the standard deviation of the noise is given by $\frac{\sigma^2}{r_n}$, where $\sigma^2 = \frac{1}{\beta}$. Therefore, the likelihood is given by:

$$p(t_n|x_n, \mathbf{w}) \sim \mathcal{N}(t_n|\mathbf{w}^T\phi(x_n), (\beta r_n)^{-1})$$

The prior is given by **(check this)**:

$$p(\mathbf{w}|\alpha) \sim \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

1.2 ML and MAP Estimation

The likelihood is given by:

$$p(\mathbf{t}|\mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\phi(x_n), (\beta r_n)^{-1})$$

Here, we assume that the t_is are independent. The above expression simplifies to:

$$\begin{aligned} p(\mathbf{t}|\mathbf{w}, \beta) &= \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\phi(x_n), (\beta r_n)^{-1}) && (1) \\ &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi(\beta r_n)^{-1}}} \exp\left(-\frac{1}{2}(\beta r_n)(t_n - \mathbf{w}^T\phi(x_n))^2\right) && (2) \end{aligned}$$

For the MAP estimate, we consider:

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)$$

Therefore, we have:

$$\begin{aligned}
p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) &\propto p(\mathbf{t}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha) \\
&= \left(\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi(\beta r_n)^{-1}}} \exp\left(-\frac{1}{2}(\beta r_n)(t_n - \mathbf{w}^T\phi(x_n))^2\right)\right) \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left(-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\right)
\end{aligned}$$

1.3 Error Function and Solution

To maximise the likelihood, we take log on both sides of 2:

$$\begin{aligned}
\log p(\mathbf{t}|\mathbf{w}, \beta) &= \sum_{n=1}^{N} \log \frac{1}{\sqrt{2\pi(\beta r_n)^{-1}}} \exp\left(-\frac{1}{2}(\beta r_n)(t_n - \mathbf{w}^T\phi(x_n))^2\right) \\
&= \sum_{n=1}^{N} \left(\log \frac{1}{\sqrt{2\pi(\beta r_n)^{-1}}} + \log\exp\left(-\frac{1}{2}(\beta r_n)(t_n - \mathbf{w}^T\phi(x_n))^2\right)\right) \\
&= \sum_{n=1}^{N} \left(\log \frac{1}{\sqrt{2\pi(\beta r_n)^{-1}}} - \frac{1}{2}(\beta r_n)(t_n - \mathbf{w}^T\phi(x_n))^2\right) \\
&= -\frac{\beta}{2}\sum_{n=1}^{N} r_n(t_n - \mathbf{w}^T\phi(x_n))^2 + \sum_{n=1}^{N} \log \frac{1}{\sqrt{2\pi(\beta r_n)^{-1}}} \\
&= -\beta E_D(\mathbf{w}) + \sum_{n=1}^{N} \log \frac{1}{\sqrt{2\pi(\beta r_n)^{-1}}}
\end{aligned}$$

Therefore, we have to minimise the error function:

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} r_n(t_n - \mathbf{w}^T\phi(x_n))^2$$

To do this, we rewrite E_D as $(T - \Phi\mathbf{w})^T R(T - \Phi\mathbf{w})$, where:

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix} \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} \mathbf{R} = \begin{pmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_N \end{pmatrix}$$

Therefore, we have:

$$\begin{aligned}
\frac{\partial E_D}{\partial \mathbf{w}} &= \frac{\partial}{\partial \mathbf{w}}(T - \Phi\mathbf{w})^T R(T - \Phi\mathbf{w}) \\
&= -2(T - \Phi\mathbf{w})^T R\Phi \\
&= 0 \\
\implies \Phi^T R(T - \Phi\mathbf{w}) &= 0 \\
\implies \Phi^T RT &= \Phi^T R\Phi\mathbf{w} \\
\implies \mathbf{w} &= (\Phi^T R\Phi)^{-1}\Phi^T RT
\end{aligned}$$