Aditya Shankar Garg
(2019ME10770)

Date :
Page No. :

Q1

for the parameter $\overline{w}^*$ such that $E_D(\overline{w})$ is minimised we need to set up the derivative of $E_D(\overline{w}) = 0$ (wrt $\overline{w}$)

$\frac{\partial}{\partial \overline{w}}(E_D(\overline{w})) = \frac{1}{2} \cdot 2 \sum r_n (y_n - w^T\phi(\overline{x}_n))\, \phi(\overline{x}_n) = 0$ { using the chain rule of derivative and using $\phi^T w = w^T \phi$

$\Rightarrow \sum_{n=1}^{N} r_n y_n \phi(x_n) = \left(\sum_{n=1}^{n} r_n \phi(\overline{x}_n)\phi(\overline{x}_n)^T\right) w$

$\Rightarrow \overline{w} = \left(\sum_{n=1}^{N} r_n \phi(\overline{x}_n)\phi(\overline{x}_n)^T\right)^{-1}\left(\sum_{n=1}^{N} r_n t_n \phi(\overline{x}_n)\right)$

we can also write the error function using matrices.

$\longrightarrow \frac{1}{2}\sum r_n (y_n - w^T\phi(x_n))^2 = \frac{1}{2}(\phi w - y)^T \underline{R}(\phi w - y)$

design matrix, Param weights. weight matrix

$\Rightarrow \frac{1}{2}(w^T\phi^T R \phi w - w^T\phi^T R y - y^T R \phi w + y^T R y)$ similar

$\Rightarrow \frac{1}{2}(w^T\phi^T R\phi w - 2y^T R\phi w + y^T R y)$

$\Rightarrow \nabla_w E_D(w) = (\phi^T R \phi w - y^T R \phi) = w^* = (A)(y^T R\phi)$

$A = (\phi^T R \phi)^{-1}$

$\Rightarrow$ in the last expression as the matrix $y^T R\phi = \phi^T R y$ (R is diagonal) $(r_1, r_2 \cdots r_n)$

$\therefore$ we can write this solⁿ as $w^* = (\phi^T R\phi)^{-1}\phi^T R y$

much more simpler and elegant form.

↳ objective 1 $\longrightarrow$ explain weighted least square in terms of data dependent noise.

$\longrightarrow$

$y_n = w^T\phi(x_n) + \epsilon_n$ some gaussian noise $\sim N(0, \sigma_n^2)$

then if we try to maximise the log-likelihood of the output by setting $\sigma_n^2 = \frac{1}{2r_n}$

Date:
Page No.:

objective 2 →

if we have r_n — copies of n^{th} data point then they will get coupled into a single term with an associated weight.

for ex → $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \begin{pmatrix} x_3 \\ y_3 \end{pmatrix}$ are repeated data points then when we compute the loss func^n then

$$E_D = (\cdot) + \frac{3}{2}\left(y_n - w^T \phi(x_n)\right)^2$$

loss due to others

associated r_n ✓.