Machine Learning week 4

Linear Regression

Kieu Son Tung

5. Oktober 2021

1 Maximizing the posterior to find w

We have that

$$p(w|x,t,\alpha,\beta) = \frac{p(t|x,w,\beta)p(w|\alpha)}{p(x,t,\alpha,\beta)} \tag{1}$$

We are trying to maximize the posterior to find w, which means we have to maximize $p(t|x,w,\beta)p(w|\alpha)$. Suppose $p(w|\alpha)$ is a normal distribution, we have

$$\begin{aligned} p(w|\alpha) &= \mathcal{N}(w|0,\alpha^{-1}I) \\ &= (\frac{\alpha}{2\pi})^{(M+1)/2} exp(-\frac{\alpha}{2}w^T w) \end{aligned} \tag{2}$$

So

$$p(w|x,t,\alpha,\beta) \propto p(t|x,w,\beta)p(w|a) \propto exp\left(-\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n,w)-t_n)^2 - \frac{\alpha}{2}w^T w\right) \tag{3}$$

We find that the maximum of the posterior is given by the minimum of

$$\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n,w)-t_n)^2 + \frac{\alpha}{2}w^T w$$

Or we minimize

$$Q = \|Xw - t\|^2 + \lambda w^T w$$

We have

$$
\begin{aligned}
\frac{\partial Q}{\partial w} &= \frac{\partial}{\partial w}(Xw - t)^T(Xw - t) + 2\lambda Iw \\
&= \frac{\partial}{\partial w}(w^T X^T - t^T)(Xw - t) + 2\lambda Iw \\
&= \frac{\partial}{\partial w}(w^T X^T Xw - w^T X^T t - t^T Xw + t^T t) + 2\lambda Iw \\
&= \frac{\partial}{\partial w}(w^T X^T Xw - 2t^T Xw + t^T t) + 2\lambda w \\
&= X^T Xw + X^T Xw - 2X^T t + 2\lambda Iw = 0 \\
\implies X^T t &= w(X^T X + \lambda I) \\
\implies w &= \left(X^T X + \lambda I\right)^{-1} X^T t
\end{aligned}
\tag{4}
$$