Proof of Homework

Song Liu

Proof. $p(\boldsymbol{w}|D) \propto p(D|\boldsymbol{w})p(\boldsymbol{w})$, under our assumption we have

$$p(\boldsymbol{w}|D) \propto \left[\prod N_{y_i}(f(\boldsymbol{x}_i,\boldsymbol{w}),\sigma^2)\right] N_{\boldsymbol{w}}(\boldsymbol{0},\sigma_w^2\boldsymbol{I})$$

Take the logarithm, we have

$$\begin{aligned}\log p(\boldsymbol{w}|D) &= \left[\sum_{i=1}^{n} -\frac{\|y_i - f(\boldsymbol{x}_i;\boldsymbol{w})\|^2}{2\sigma^2}\right] + \frac{-\|\boldsymbol{w}\|^2}{2\sigma_{\boldsymbol{w}}^2} + \text{const.}\\ &= \left[\sum_{i=1}^{n} -\frac{\|y_i - \langle \boldsymbol{w},\boldsymbol{x}\rangle\|^2}{2\sigma^2}\right] + \frac{-\|\boldsymbol{w}\|^2}{2\sigma_{\boldsymbol{w}}^2} + \text{const.}\\ &= -\frac{\boldsymbol{w}^\top\boldsymbol{w}}{2\sigma_{\boldsymbol{w}}^2} - \frac{\boldsymbol{w}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{w}}{2\sigma^2} + \frac{2\boldsymbol{w}^\top\boldsymbol{X}\boldsymbol{y}}{2\sigma^2} - \frac{\boldsymbol{y}\boldsymbol{y}^\top}{2\sigma^2} + \text{const.}\end{aligned}$$

Let us ignore all the terms that are not related to $\boldsymbol{w}$.

$$\begin{aligned}&-\frac{\boldsymbol{w}^\top\boldsymbol{w}}{2\sigma_{\boldsymbol{w}}^2} - \frac{\boldsymbol{w}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{w}}{2\sigma^2} + \frac{2\boldsymbol{w}^\top\boldsymbol{X}\boldsymbol{y}}{2\sigma^2}\\ =&-\frac{1}{2\sigma^2}\left\{\boldsymbol{w}^\top\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]\boldsymbol{w} + 2\boldsymbol{w}^\top\boldsymbol{X}\boldsymbol{y}\right\}\\ =&-\frac{1}{2\sigma^2}\left\{\boldsymbol{t}\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]\boldsymbol{t}^\top\right\} + \text{const.},\end{aligned}$$

where $\boldsymbol{t} = \boldsymbol{w} - \left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\boldsymbol{x}\boldsymbol{y}$.

We can see that $\log p(\boldsymbol{w}|D) = -\frac{1}{2\sigma^2}\left\{\boldsymbol{t}\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]\boldsymbol{t}^\top\right\} + \text{const.}$, so

$$p(\boldsymbol{w}|D) = N_{\boldsymbol{w}}\left(\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\boldsymbol{x}\boldsymbol{y}, \sigma^2\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\right).$$

The second part of the proof follows Bishop's book, 2.115. We have

$$p(\hat{y}|\boldsymbol{x};\boldsymbol{w}) = N_{\hat{y}}(f(\boldsymbol{x};\boldsymbol{w}),\sigma^2),$$

and

$$p(\boldsymbol{w}|D) = N_{\boldsymbol{w}}\left(\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\boldsymbol{x}\boldsymbol{y}, \sigma^2\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\right),$$

2.115 tells us

$$p(\hat{y}|\boldsymbol{x},\boldsymbol{w}) = N_{\hat{y}}\left[\boldsymbol{x}^\top\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\boldsymbol{x}\boldsymbol{y}, \sigma^2 + \sigma^2\boldsymbol{x}^\top\left[\frac{\sigma^2}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\boldsymbol{x}\right].$$

$\square$