8. MSE with Probabilistic Perspective

MSE with Probabilistic Perspective

회귀에서 사용되는 MSE 손실함수를 최소로 하는 파라미터는 MLE 관점에서 데이터를 가장 잘 나타내는 파라미터이다.

회귀의 MSE를 사용할 때는 출력값이 가우시안 분포를 따른다고 가정한다.

Gaussian PDF

$$p(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\log p(x;\mu,\sigma) = -\log \sigma\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

$$-\log p(x;\mu,\sigma) = \log \sigma\sqrt{2\pi} + \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

Get gradient of NLL

log likelihood에 가우시안 pdf를 넣어서 계산해본다. sigma에 관심이 없다 생각하고 평균인 mu에만 관심을 가지면 최종적으로 NLL의 식에서 MSE가 나오는 것을 볼 수 있다.

$$-\log p(y_i|x_i;\phi,\psi) = \log \sigma_\psi(x_i)\sqrt{2\pi} + \frac{1}{2}\left(\frac{y_i-\mu_\phi(x_i)}{\sigma_\psi(x_i)}\right)^2, \text{ where } \theta = \{\phi,\psi\}.$$

$$\begin{aligned}
-\nabla_\phi \log p(y_i|x_i;\phi,\psi) &= \nabla_\phi \log \sigma_\psi(x_i)\sqrt{2\pi} + \nabla_\phi \frac{1}{2}\left(\frac{y_i-\mu_\phi(x_i)}{\sigma_\psi(x_i)}\right)^2 \\
&= \frac{1}{2\cdot\sigma_\psi(x_i)^2}\nabla_\phi\big(y_i-\mu_\phi(x_i)\big)^2 \\
&= \alpha\cdot\nabla_\phi\big(y_i-\mu_\phi(x_i)\big)^2, \text{ where } \alpha = \frac{1}{2\cdot\sigma_\psi(x_i)^2}.
\end{aligned}$$

μ 에만 관심이 있음

$\Leftrightarrow y_i - \hat{y}_i$